Neural dynamics of the attentional blink revealed by encoding orientation selectivity during rapid visual presentation

Matthew F. Tang[1,2], Lucy Ford[1], Ehsan Arabzadeh[2,3], James T. Enns[4,5], Troy A.W. Visser[5], & Jason B. Mattingley[1,2,6,7]

1. Queensland Brain Institute, The University of Queensland, Brisbane, QLD, Australia
2. Australian Research Council Centre of Excellence for Integrative Brain Function, Victoria, Australia
3. Eccles Institute of Neuroscience, John Curtin School of Medical Research, The Australian National University, Canberra, ACT, Australia
4. Department of Psychology, The University of British Columbia, Vancouver BC, Canada
5. School of Psychological Sciences, The University of Western Australia, Perth, WA, Australia
6. School of Psychology, The University of Queensland, Brisbane, QLD, Australia
7. Canadian Institute for Advanced Research (CIFAR)

Corresponding author

Matthew F. Tang

Queensland Brain Institute

The University of Queensland

St Lucia, QLD, Australia

Email: m.tang1@uq.edu.au

Abstract

The human brain is inherently limited in the information it can make consciously accessible. When people monitor a rapid stream of visual items for two targets, they can typically report the first, but not the second target, if these appear within 200-500 ms of each other, a phenomenon known as the attentional blink (AB). No work has determined the neural basis for the AB, partly because conventional neuroimaging approaches lack the temporal resolution to adequately characterise the neural activity elicited by each item in a rapid stream. Here we introduce a new approach that can identify the precise effect of the AB on behaviour and neural activity. Specifically, we employed a multivariate encoding approach to extract feature-selective information carried by randomly-oriented gratings within a rapid serial stream. We show that feature selectivity is enhanced for correctly reported targets and suppressed when the same items are missed. By contrast, no such effects were apparent for irrelevant distractor items. Our findings point to a new theoretical account that involves both short- and long-range temporal interactions between visual items competing for consciousness.

Introduction

Despite the remarkable capacity of the human brain, it is found wanting when undertaking multiple tasks concurrently, or when several goal-relevant items must be dealt with in rapid succession. These limitations are particularly evident when individuals are required to execute responses to multiple items under time pressure[1,2], or when they must report relevant target items that appear briefly and in rapid succession[3-5]. Elucidating the source of these limitations has been a persistently difficult challenge in neuroscience and psychology. While the neural bases for these processing limits are not fully understood, it is widely assumed that they are adaptive because they provide a mechanism by which selected sensory events can gain exclusive control over the motor systems responsible for goal-directed action.

Here we address a long-standing question concerning the neural basis of the widely-studied 'attentional blink' (AB), where observers often fail to report the second of two target items (referred to as T2) when presented within 200-500 ms of the first target (T1) in a rapid stream of distractors[3-5]. Functional magnetic resonance imaging (fMRI) lacks the temporal resolution to accurately characterise neural activity associated with the rapid serial visual presentation (RSVP) tasks presented at rates of 8-12 Hz, which are commonly used to elicit the AB[6,7]. Even electroencephalography (EEG), which has relatively good temporal resolution, produces smeared responses to items in an RSVP stream[8]. Furthermore, mass-univariate approaches applied to fMRI or EEG data only measure overall neural activity while providing no information about how neural activity represents featural information carried by single items (e.g., their orientation).

Here we overcame these limitations by combining recently-developed multivariate modelling techniques for neuroimaging[9-16] with a novel RSVP task designed to determine the neural and behavioural basis for the AB. Forward (or inverted) encoding modelling determines the neural representation of feature-selective information contained within patterns of brain activity, using multivariate linear regression[9-16]. This approach allowed us to explicitly measure the neural representation of specific features – in this case, orientation-selective information elicited by grating stimuli – separately for each item within an entire RSVP stream.

We used this approach to address two central theoretical questions. First, does selection of a target from within an RSVP stream increase the gain or the precision of its neural representation? Previous efforts to answer this question in the domain of spatial attention have come from single cell recordings in non-human primates[17,18], as well as whole-brain activity measured using fMRI[15,19] and EEG[9] in humans. With few exceptions[20], these studies have found that spatial attention increases the gain of feature-selective processing of attended items. By contrast, feature-based manipulations of attention, in which specific characteristics of an item such as its colour or motion are cued for selective report, typically result in a sharpening of neural selectivity[20-22]. To date, it remains unknown whether the limits of temporal attention in the AB are associated with changes in neural tuning to targets, distractors, or both classes of items. The neural response in human primary visual cortex[6] and macaque lateral intraparietal area[23] to the second target is reduced overall on AB trials compared with non-AB trials, while subtraction-based EEG designs have shown that a late-stage

component of the ERP (the N400) is reduced 200-400 ms after target presentation[8]. Critically, however, these measures cannot determine how the AB affects the neural representation of visual information, which could conceivably reflect a reduction in gain, an increase in tuning sharpness, or both.

A second, unresolved theoretical question concerns the source of the AB. Existing theories have often attributed the AB to either extended processing of the first target, or to inadvertent distractor processing. In the first class of theories, it is assumed that all items generate representations in early visual areas, but that the system inhibits items after T1 detection to avoid contamination by distractors[4,24-27]. On other accounts (so-called 'distractor-based' theories), the AB is assumed to reflect a cost associated with switching between target and distractor processing[24]. Finally, a third class of theories argues that the representation of the second target can become merged with either the first target or the distractors[25,26]. This class of theories is motivated by the finding that the perceived order of targets is often reversed (i.e., T2 is reported as appearing before T1).

Our RSVP task consisted of a stream of randomly-oriented Gabor gratings, with two higher-spatial frequency targets set amongst lower-spatial frequency distractors (Figure 1A and Movie 1). At the end of the stream, participants were asked to reproduce the orientations of the two targets (Figure 1B). Critically, the orientation of each item in the stream was uncorrelated with the orientation of all other items (Figure 1C), thus permitting the use of linear regression analyses to separately extract the influence of each item in the stream on neural activity measured by EEG, and on behavioural

112 reports of the orientations of the two targets. These aspects of the experimental design
113 allowed us to quantify the influence of both targets and distractors on participants'
114 perceptual reports and on their associated neural representations.
115

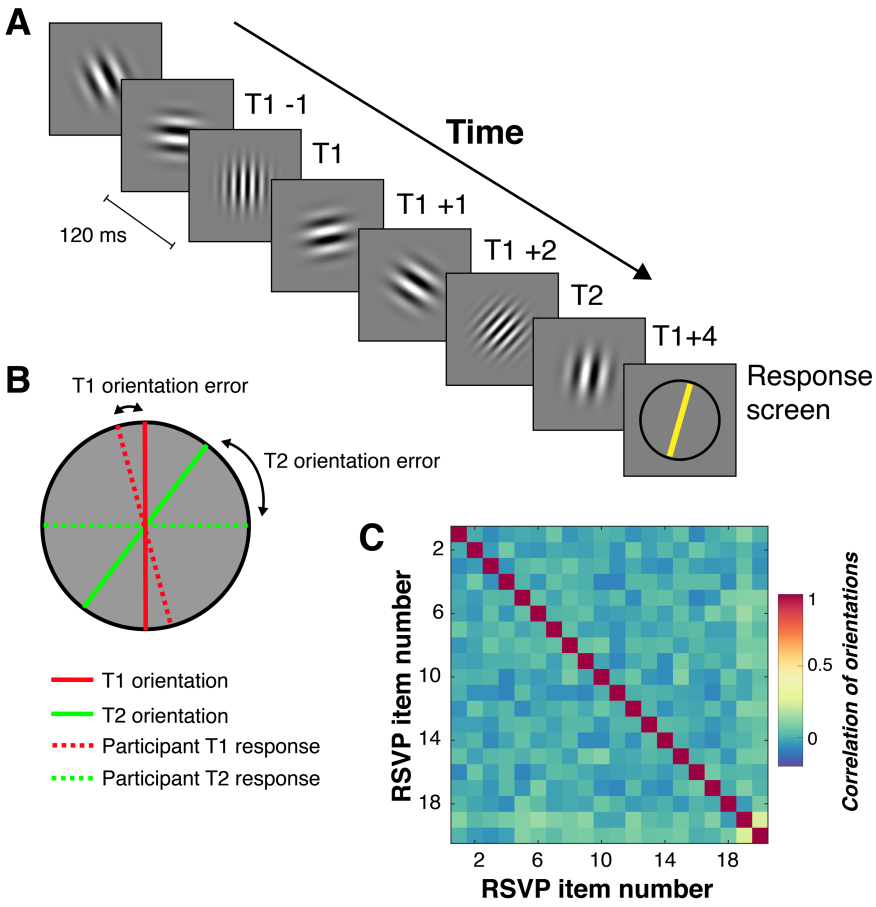


Figure 1. Schematic of stimuli and timing of displays in the novel rapid serial visual presentation (RSVP) task. **A.** An illustration of a typical trial in the RSVP task, which consisted of 20 sequentially presented Gabor patches at fixation. Each of the twenty items within a single RSVP stream was presented for 40 ms, with an 80 ms blank interval between items (120 ms inter-stimulus interval), yielding an 8.33 Hz presentation rate. The number of items (Lag) between the first (T1) and second (T2) targets was varied to measure the temporal duration of the AB. At the end of each RSVP stream, participants reproduced the orientations of T1 and T2 (higher spatial-frequency gratings) in the order in which they were presented by adjusting an on-screen cursor at the end of the trial. They were asked to determine the orientations as accurately as possible and were not given any time restriction to do this. Visual feedback was provided following the response. **B.** A schematic of the feedback screen for responses. **C.** The correlation values between orientations of the RSVP items over trials in Experiment 1. As Gabor orientations were randomly drawn (without replacement) on each trial, across all trials the orientation of any given item in the stream was uncorrelated with the orientation of any other item. This permitted the use of regression-based approaches to isolate the behavioural and neural processing of individual items independently of surrounding items within the stream. The correlations were calculated for each participant and are displayed as averaged across participants.

Movie 1. Examples of two trials from the RSVP task.

To preview the results, behavioural target reproduction replicated the key hallmarks of the AB effect: the orientation of T1 was reported with a relatively high degree of accuracy, whereas orientation judgements for T2 were degraded when T2 appeared 200-400 ms after T1. Forward encoding analyses of EEG activity showed that targets evoked greater orientation-selective information than distractors when T2 was accurately reported (i.e., in non-AB trials), and that orientation information evoked by both targets was suppressed, relative to the distractors, when T2 was missed (i.e., in AB trials). Critical to our first question of whether focused attention influences the gain or precision of feature-specific representations, only the gain of the encoded EEG response was affected by T2 response accuracy.

With respect to our second question — whether accuracy in registering the second target is linked to the processing of T1 or to the intervening distractors — the evidence was in favour of T1-based theories of the AB. We found no evidence to suggest that neural representations of the distractors were affected by target

153 processing. Finally, we describe a novel observation – one not predicted by any theory
154 of the AB – namely, a significant interaction of the specific features of T1 and T2,
155 implying a previously unknown long-range temporal integration of target representations
156 within rapid sequential visual streams.
157
158 **Results**
159 **Experiment 1 - Behavioural measurement of target orientation perception for**
160 **items in RSVP streams**
161 ***Behavioural results replicate hallmarks of the attentional blink***
162 Participants' (N=22) response errors (i.e., the difference between the presented
163 and reported orientation for each target) were centred around 0°, verifying that they
164 were able to perform the task as instructed. Figure 2A captures the temporal dynamics
165 of the AB, such that accuracy was affected by target position (T1 or T2) and Lag.
166 Specifically, at Lag 1 accuracy for both T1 and T2 was degraded relative to accuracy at
167 the other lags (2, 3, 5 & 7). Moreover, at Lags 2 and 3, T1 accuracy was high whereas
168 T2 accuracy was relatively poor. This was largely due to an increase in the baseline
169 guessing rates (where errors occurred evenly across all orientations). Finally, at longer
170 temporal separations (Lags 5 and 7), target accuracy was similar for both items.

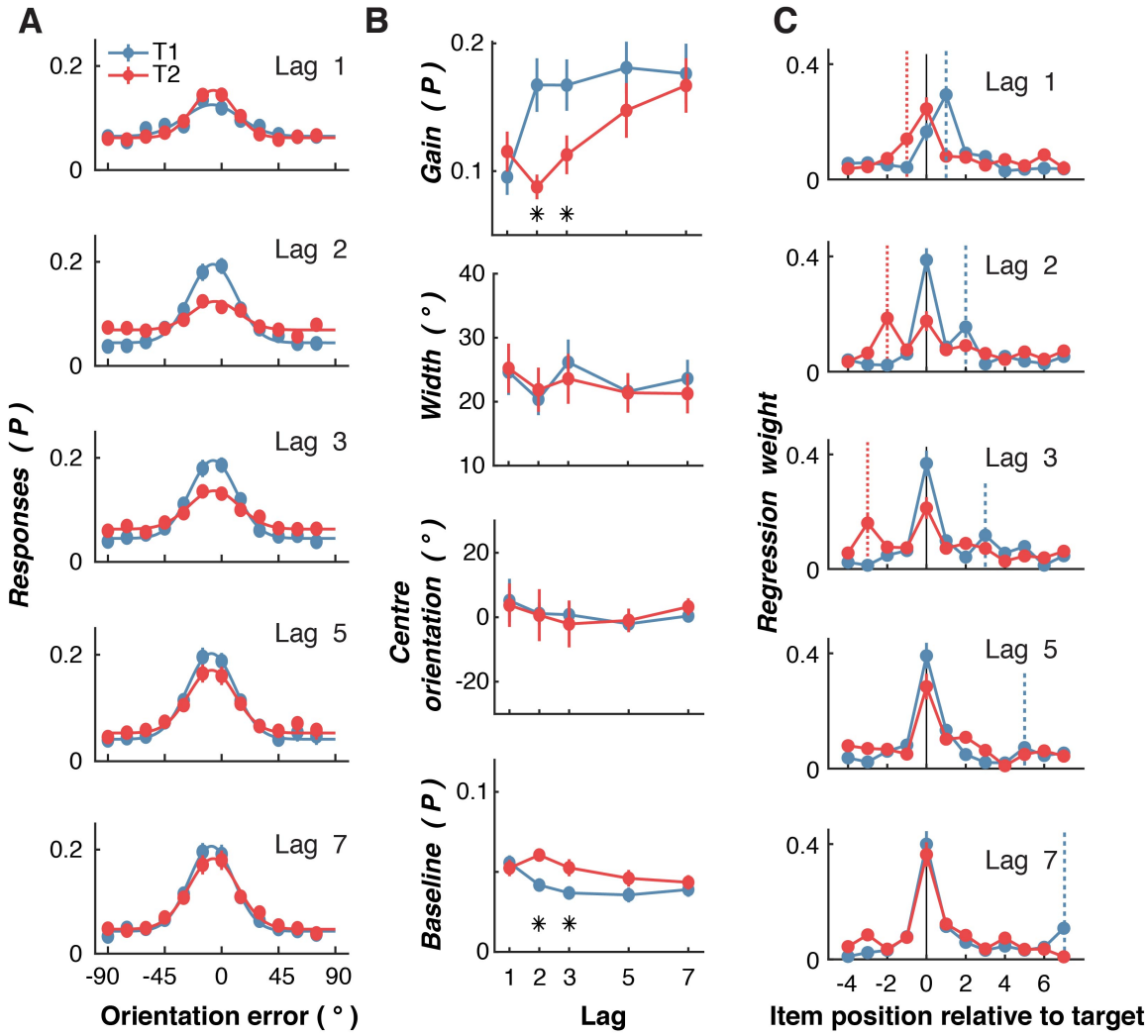


Figure 2. Behavioural results for the RSVP task in Experiment 1. **A.** The distribution of response errors (difference between presented and reported orientation) across participants (N=22) for the first (T1) and second (T2) target for each Lag condition. The line shows fitted four-parameter Gaussian function. **B.** Quantified behavioural responses for the four parameters of the fitted Gaussian function (see Supplementary Figure 1) for each participant. Gain shows the amplitude, width shows the standard deviation of the function, centre orientation is the mean (which should be centred around 0° for unbiased estimates), and baseline is a constant parameter accounting for non-orientation selective responses which indicates guessing. Asterisks indicate Bonferroni-corrected significant differences at $p < .05$. **C.** Regression results for the influence of distractors and targets on participants' responses. Higher regression weights indicate that a given item's orientation was more influential for determining the reported orientation. The dotted vertical lines indicate the position of the other target (colour matched). Consider, for example, the panel depicting Lag 2 results. For T1 report, T2 occurred at item+2 as indicated by the dotted blue line, whereas when for T2 report, T1 occurred at item-2, as indicated by the dotted red line. Across all panels, error bars indicate ∓1 standard error of mean.

Gain or precision of behavioural response?

We fitted Gaussian functions to each individual's data to quantify how the AB affected target perception (Figure 2B; see Methods and Supplementary Figure 1). The accuracy reduction for T2 at Lags 2 and 3 was primarily linked to a reduction in gain. A 2 (Target; T1,T2) x 5 (Lag; 1,2,3,5,7) within-subjects ANOVA showed the gain parameter was affected by Target ($F(1,21)=10.00$, $p=0.005$, $\eta_p^2=0.32$) and Lag ($F(4,84)=11.66$, $p<0.0001$, $\eta_p^2=0.36$), and the interaction between these factors ($F(4,84)=7.10$, $p<0.0001$, $\eta_p^2=0.25$). Critically for our first theoretical question, the spread (width) of orientation errors was unaffected by the factors of Target ($F(1,21)=0.10$, $p=0.76$, $\eta_p^2=0.005$) or Lag ($F(4,84)=0.55$, $p=0.70$, $\eta_p^2=0.03$), or by the interaction between these factors ($F(4,84)=0.19$, $p=0.94$, $\eta_p^2=0.01$). The baseline parameter, which reflects guessing of random orientations, was also significantly affected by the factors of Target ($F(1,21)=12.72$, $p=0.002$, $\eta_p^2=0.38$) and Lag ($F(4,84)=4.82$, $p=0.002$, $\eta_p^2=0.19$), and by the interaction between them ($F(4,84)=5.04$, $p=0.001$, $\eta_p^2=0.19$). These same effects were also evident when the data were not normalized (Supplementary Figure 2), and with a wide range of parameters to specify the orientation errors (Supplementary Figure 3).

Taken together, these results are consistent with a previous AB study using similar analysis methods[27]. They also lend weight to the global workspace theory of consciousness in the AB[28], which argues that participants either see the target and have full awareness of it (allowing them to respond precisely), or they have no awareness (and so simply guess randomly). By contrast, the results are inconsistent with the

opposing view that the AB involves a noisier (i.e., weaker precision) signal for the target that is inaccurately reported[29].

Targets, not distractors, influence judgements of T1 and T2 orientation

To evaluate the influence of distractors on participants' reports, we aligned the orientations of the items relative to target position within the RSVP stream (-4 to +7 items) and constructed a regression matrix to predict the behavioural response for each target. If the orientation of an item is influential in determining the reported orientation, the regression weight will be relatively high (Figure 2C). As expected, for all lags, each target item was influenced principally by its own orientation. The one exception was the item at Lag 1, where the reported orientation of T1 was as strongly influenced by the orientation of T2 as by the orientation of T1. This observation is in line with numerous studies which have suggested that temporal order information can be lost for consecutive targets[25,26,30]. This phenomenon, also known as *Lag 1 switching,* where the perceived order of the targets is reversed, explains why the accuracy of orientation judgements on both T1 and T2 was reduced at Lag 1 (see also Supplementary Figure 4). By contrast, for items at Lags 2 and 3, orientation judgements on T1 were only marginally influenced by the orientation of T2 (i.e., for items at positions +2 and +3, respectively, in the RSVP stream). However, at these same lags (where the AB was maximal) T2 reports were significantly influenced by T1 orientation (i.e., for items at positions -2 and -3, respectively). Importantly, there was no reliable influence of distractors on reported target orientation at any lag, suggesting distractors played little or no role in target orientation errors.

234 ***Long-range temporal integration of target orientations***

235 One account of the AB[25,26] has suggested that successive targets presented at
236 short lags are integrated into a single 'episodic trace', which accounts for Lag 1
237 switching. With the present task, we can directly quantify how targets are integrated by
238 looking for systematic biases in the reported orientation of a given target based on its
239 orientation difference with respect to the other target. Figure 3A shows orientation
240 judgement errors as a function of the difference between the two target orientations.
241 While the average orientation error is centred on 0°, the perceived orientation of either
242 target (T1 or T2) was significantly biased toward the orientation of the other target within
243 the RSVP stream at early Lags. Furthermore, these biases were orientation-tuned, such
244 that the largest bias occurred when targets differed by approximately 45°, somewhat
245 analogous to serial dependency effects[31-33]. This profile of biases suggests response
246 integration, rather than replacement, as the latter would predict that only the orientation
247 of T2 should drive the reported orientation of T1. Instead, and consistent with our linear
248 regression analysis (see Figure 2C), the bias reflected the *difference* between target
249 orientations, which supports the idea that the critical features of the two targets are
250 assimilated over time[25,26].
251

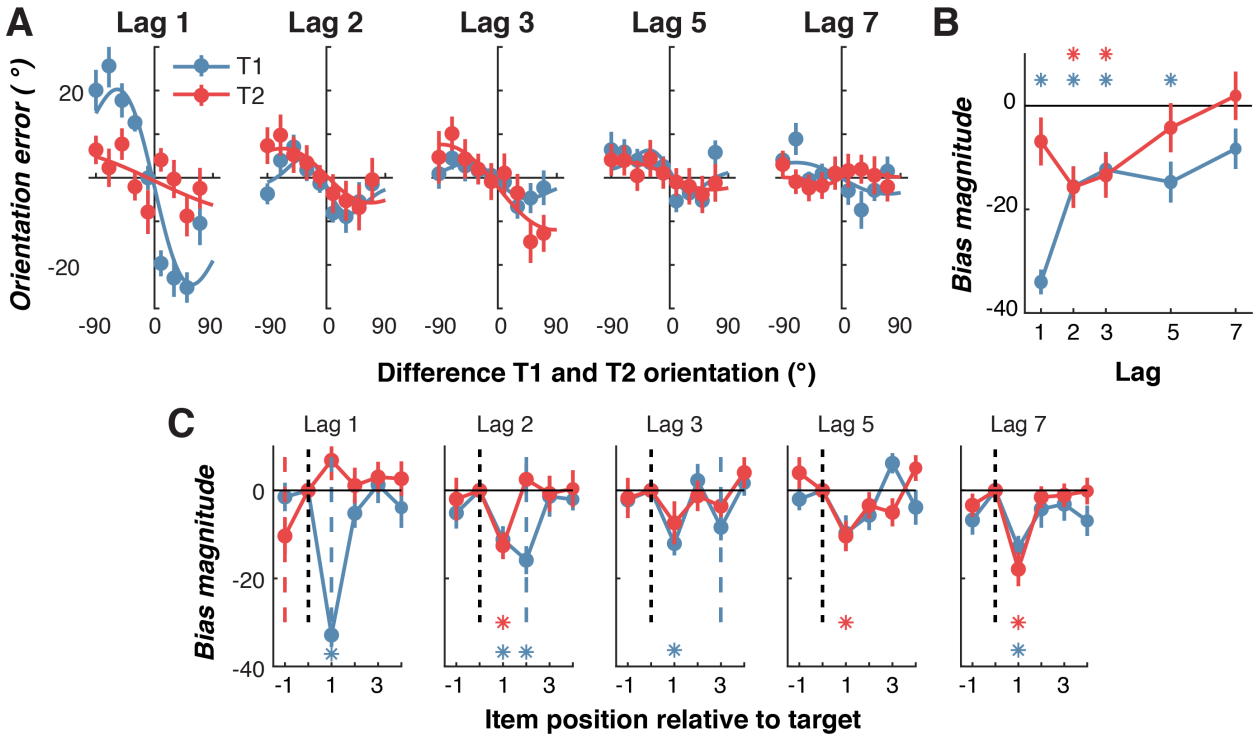


Figure 3. The influence of targets and distractors on the reported orientation in Experiment 1. **A.** Orientation error (the difference between presented and reported orientation) plotted against the difference between T1 and T2 orientations (divided into 30° bins, for clarity of presentation). Negative values on the X-axis indicate that a given target was rotated clockwise relative to the other target. For instance, when examining T1, a negative value indicates that T2 was rotated clockwise relative to T1, whereas a positive value indicates that T2 was rotated anti-clockwise relative to T1. For T1, the plotted values reflect the calculation of T1 minus T2, and vice versa for the calculation of T2, to ensure values were equivalent for the comparison of interest. The same convention applies to orientation error, shown on the Y-axis. The fitted line is the first derivative of a Gaussian (D1) function showing the orientation-tuned gain and width of the response. **B.** Bias magnitude was quantified across participants by fitting the D1 function to each participant's (non-binned) data, with the gain showing bias magnitude. A negative gain on the Y-axis indicates that the perceived orientation was biased toward, rather than away from, the other target. **C.** Bias magnitude by difference with target and distractors. For both T1 and T2, the difference between the target and the item was found in the same manner as in **A**. We fit the D1 function to find the magnitude of bias induced by each item for each participant. The dotted coloured lines indicate the temporal position of the other target (T1 = blue, T2 = red). For all panels, the asterisks indicate, for each target, at which lags the bias was significantly greater than zero (Bonferroni-corrected one-sample t-test $p<0.05$). Across all panels, error bars indicate ∓1 standard error of mean.

We fit first derivative of Gaussian (D1) functions[34-36] to quantify the amount of orientation-selective bias for both targets at each Lag for each participant. A 2 (Target; T1,T2) x 5 (Lag; 1,2,3,5,7) within-subjects ANOVA revealed significant main effects of Target ($F(1,21)=5.04$, $p=0.04$, $\eta_p^2=0.19$) and Lag ($F(4,84)=6.54$, $p<.0001$, $\eta_p^2=0.24$), and a significant interaction ($F(4,84)=6.14$, $p<.0001$, $\eta_p^2=0.27$). For T1 reporting, the bias was significantly greater than chance at all intervals except Lag 7, whereas for T2, there was a significant bias at Lags 2 and 3 only (Bonferroni-corrected one-sample t-test, all $ps<0.05$). As might be expected[25,26], the 'attraction' bias in target reports was strongest when the two targets were presented with no intervening distractors between them (i.e., at Lag 1). An entirely novel finding, however, is that there was an equally strong attraction bias between targets presented at Lags 2 and 3 (see Figure 3B), even though participants were not explicitly aware of the orientation of T2 on AB trials.

Only the distractor immediately following each target biases reporting

Previous work suggests that distractor processing can significantly interfere with target processing[37-39], particularly for the immediate post-target item which can be integrated into the target representation[25,26,30]. To determine whether this was the case in our data, we repeated the previous analysis but used the difference in orientation between the target and each of the other items in the RSVP stream (Figure 3C). For most lags, the reported target orientation was significantly attracted toward the immediately following distractor, but was not reliably influenced by any other distractor. A 2 (Target; T1, T2) x 5 (Lag; 1,2,3,5,7) x 4 (Item position; -1,1,2,3,4) within-subjects ANOVA confirmed a significant three-way interaction between the factors

298 ($F(16,336)=4.11, p<.0001, \eta_p^2=0.16$). At Lag 1, there was no influence of distractors on
299 reported orientations for either T1 or T2. Taken with the previous result, this suggests
300 that the representation of a given target is influenced by both the other target and by the
301 post-target item. The results suggest that when the visual system detects a target, it
302 automatically integrates features from the immediately subsequent item. This is
303 consistent with previous studies that have highlighted the importance of masking by the
304 item immediately following the target in eliciting the AB[40].
305
306 **Experiment 2 – Characterizing neural representations of target orientation**
307 **information in RSVP streams**
308 We next characterized the neural activity elicited by individual RSVP items, and
309 determined how this was affected by the AB. In Experiment 2, a group of 23 new
310 participants undertook the RSVP task introduced in Experiment 1 while neural activity
311 was concurrently measured using EEG. The method was identical in all respects,
312 except that we now included targets only at Lags 3 and 7 (i.e., a single target inside and
313 outside the AB, respectively) to increase the within-subject power for the EEG analyses.
314 ***Behavioural results***
315 The behavioural results replicated, in all important respects, those found in
316 Experiment 1. As shown in Figure 4A, participants performed well overall, and their
317 orientation judgements for T1 and T2 were centred on the presented orientations. As in
318 Experiment 1, we fit Gaussian functions to quantify the results (Figure 4B). For the gain
319 parameter, a 2 (Target; T1, T2) x 2 (Lag; 3,7) within-subjects ANOVA revealed

significant main effects of Target ($F(1,22)=11.63$, $p=0.003$, $\eta_p^2=0.35$) and Lag ($F(1,22)=18.70$, $p<0.0001$, $\eta_p^2=0.46$), and a significant interaction ($F(1,22)=40.19$, $p<0.0001$, $\eta_p^2=0.65$). Likewise for the baseline parameter, there were significant effects of Target ($F(1,22)=8.96$, $p=0.007$, $\eta_p^2=0.30$) and Lag ($F(1,22)=12.21$, $p=0.002$, $\eta_p^2=0.36$), and a significant interaction ($F(1,22)=7.91$, $p=0.01$, $\eta_p^2=0.26$). By contrast, there were no significant main effects and no interaction for the width parameter (Target ($F(1,22)=1.19$, $p=0.29$, $\eta_p^2=0.05$; Lag ($F(1,22)=3.90$, $p=0.06$, $\eta_p^2=0.15$); interaction ($F(1,22)=0.14$, $p=0.71$, $\eta_p^2=0.006$).

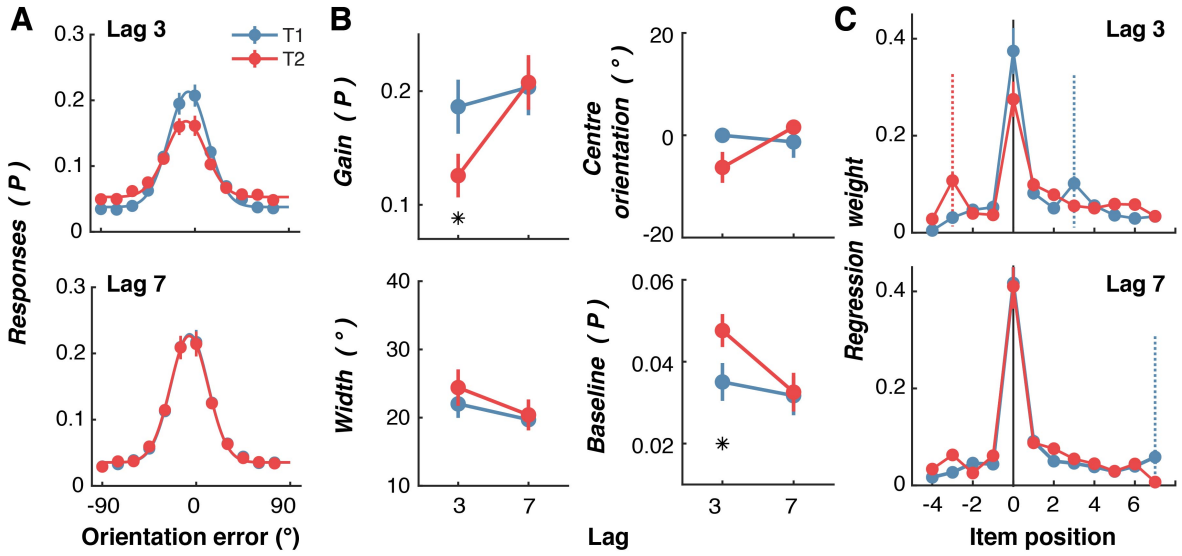


Figure 4. Behavioural results for the RSVP task in Experiment 2. **A.** Aggregate response accuracy across participants (difference between presented and reported orientations) for T1 and T2, shown separately for Lag 3 and Lag 7 trials. Lines are fitted Gaussian functions. **B.** Quantified behavioural responses for the four parameters of the fitted Gaussian functions (gain, width, centre orientation and baseline) to each participant's data. Asterisks indicate significant differences at Bonferroni-corrected $p<0.05$. **C.** Regression results for the influence of distractors and targets on participants' responses. The dotted vertical lines indicate the position of the other target (colour matched). Consider, for example, the panel depicting Lag 3 results. For T1 report, T2 occurred at item+3 as indicated by the dotted blue line, whereas when for T2 report, T1 occurred at item-3, as indicated by the dotted red line. Across all panels error bars indicate ∓1 standard error of mean.

Orientation selectivity of RSVP items

We next applied forward modelling to the EEG data recorded during the task to quantify orientation information contained within multivariate patterns of neural activity. Because the orientations of successive items were uncorrelated, we were able to quantify orientation selectivity for each grating without contamination from adjacent items. Forward encoding uses a linear regression-based approach to find multivariate patterns of EEG activity that are selective for features of interest – in this case orientation. As no previous study has used forward encoding in conjunction with rapid visual presentations, we first verified that orientation selectivity for each of the 20 RSVP items could be extracted separately using this approach, and at what time point any such response was evident. To do this, we constructed 20 encoding models, one for each of the item position for the 20-item RSVP stream, based on the orientations presented for that item across trials.

As shown in Figure 5, the forward encoding revealed robust and reliable feature selectivity derived from patterns of EEG activity for each of the gratings presented during the RSVP. Each item's orientation was successfully decoded over a time window that extended from 74 to 398 ms after the item was presented. Examination of the neural responses to each of the 20 items within the RSVP stream (Figure 5C) shows that feature selectivity was evident as a series of regularly spaced, short-lived impulse responses, each with a delay of around 50 ms from grating onset and lasting approximately 300 ms. To quantify these observations, we fit Gaussian functions to the forward encoding results for each item separately for each participant and at each time

363 point. There was significant feature selectivity (given by the gain of the Gaussian) for
364 each item immediately after it was presented but not before (Figure 5D). These
365 representations were temporally overlapping, such that multiple orientation-selective
366 responses (~3) were detectable at the same time. Taken together, the forward encoding
367 analyses verify that it is possible to reliably recover the presented orientation of every
368 RSVP item from the multivariate pattern of neural activity recorded using EEG.
369

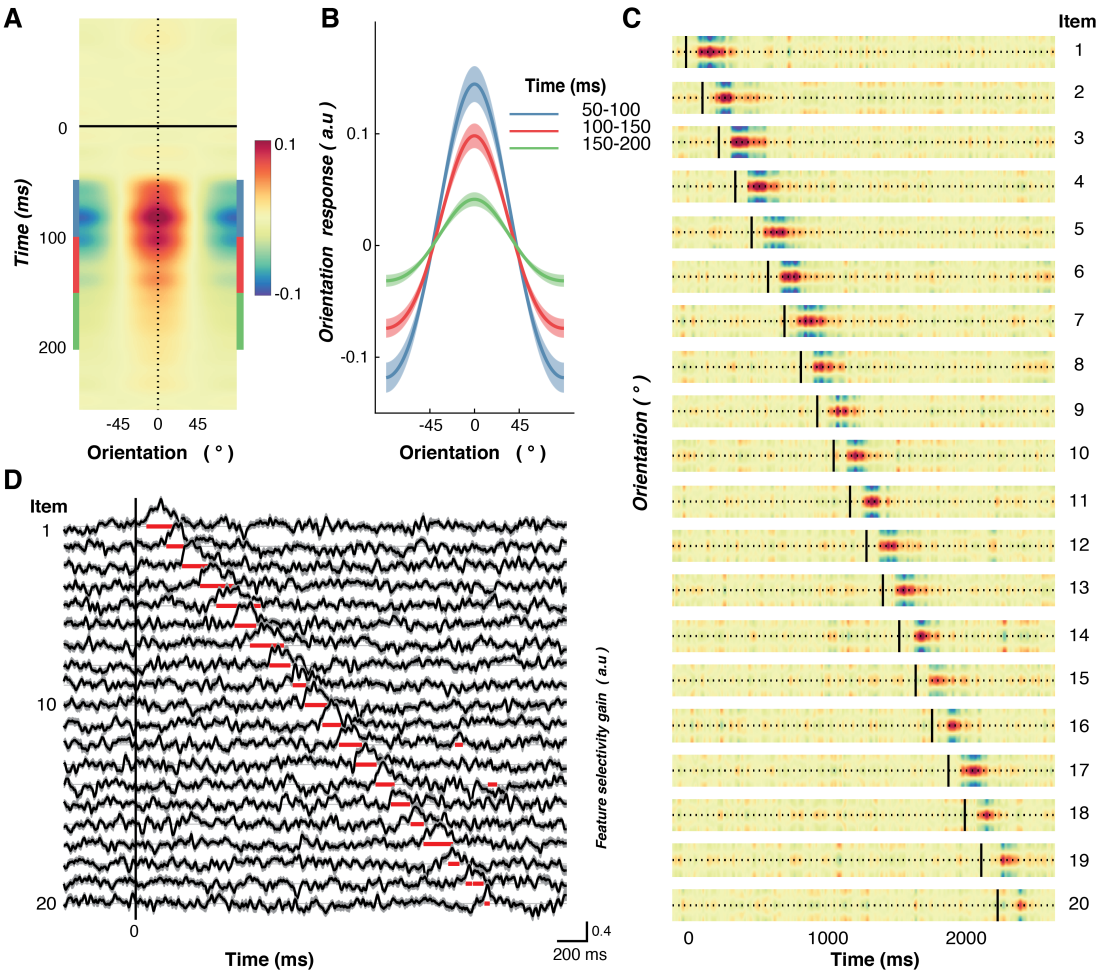


Figure 5. Feature (orientation) selectivity for RSVP items derived using forward encoding modelling of multi-variate patterns of EEG activity in Experiment 2. **A**. Forward encoding results aligned at the time of item onset and the presented orientation across all participants. All representations have been re-aligned so that the presented orientation is equivalent to 0°. **B.** Forward encoding results averaged over 50 ms bins (shown by corresponding colour in (A) following each item. Feature selectivity peaks around 50-120 ms after the onset of each item and persists for ~200 ms. **C.** Forward encoding results for each item in the RSVP stream. Vertical black lines indicate the presentation time of each of the 20 items within the RSVP stream. The dotted horizontal line indicates the presented orientation. The colour scale is the same as in panel A. **D.** Gaussian distributions were fitted to each participant's data for each item in the stream, with the gain showing feature selectivity. The red horizontal line segments underneath each trace indicate timepoints over which feature selectivity was significantly different from zero (i.e., where feature selectivity was greater than what would be expected by chance; two-tailed, sign-flipping cluster-permutation, alpha $p<0.05$, cluster alpha $p<0.05$, N permutations = 20,000), which occurs immediately following item presentation. Across all panels shading indicates ∓ 1 standard error of the mean across participants. a.u = arbitrary units.

The attentional blink is associated with reduced feature-selective information for the second target

We next examined how neural representations of the target items were affected by the AB. To increase signal-to-noise for training the encoding model, we aligned the EEG data to the presentation time of each item in the RSVP stream and applied the same forward encoding procedure. This meant that the model was trained and tested across 12,000 presentations (600 trials by 20 RSVP items; see Figure 6). To determine the effect of the AB on orientation-selectivity, we separated the forward encoding results by target (T1,T2) and T2 accuracy (correct, incorrect). For the purposes of the analyses, trials were scored as correct if the reported orientation was within ±30 degrees of the presented orientation, a criterion which yielded roughly equal correct and incorrect trials at Lag 3. In line with the AB literature, for all the EEG analyses we only included trials where participants correctly identified T1. Applying these criteria yielded the classic AB effect (Supplementary Figure 5). A 2 (Lag; 3,7) x 2 (Target; T1,T2) within-subjects ANOVA applied to these scores revealed significant main effects of Lag ($F(1,22)=199.52$, $p<0.0001$, $\eta_p^2=0.90$) and Target ($F(1,22)=8.58$, $p=0.008$, $\eta_p^2=0.28$), and a significant interaction ($F(1,22)=9.64$, $p=0.005$, $\eta_p^2=0.31$). Follow-up t-tests showed that Lag 3 accuracy was significantly lower than Lag 7 accuracy for T2 items ($t(22)=3.15$, Bonferroni $p=0.01$, $d=0.66$) but not for T1 items ($t(22)=1.97$, Bonferroni $p=0.12$,$d=0.41$).

We again fitted Gaussians to each time point to quantify the amount of feature-selective information evoked by the targets. For both T1 and T2, there was significant

411 feature-selective activity shortly after each item appeared (Figure 6A). For Lags 3 and 7,
412 there was no difference between correct and incorrect trials for the T1 representation.
413 For T2, however, incorrect trials resulted in a significantly decreased feature-selective
414 response (cluster p=0.02) relative to correct trials shortly after each item appeared (100
415 to 150 ms) at Lag 3, although the response was not completely suppressed. There were
416 no significant differences in the orientation-selective response between correct and
417 incorrect trials for T2 at Lag 7, suggesting the suppression is caused by the AB rather
418 than general target detection. This was expected because the AB typically lasts less
419 than 500 ms, and is consistent with the current behavioural results showing an AB at
420 Lag 3 but not at Lag 7. Performing the same analysis on the other parameters of the
421 Gaussian (width, centre, baseline) showed no effect of the AB (Supplementary Figure
422 6).
423

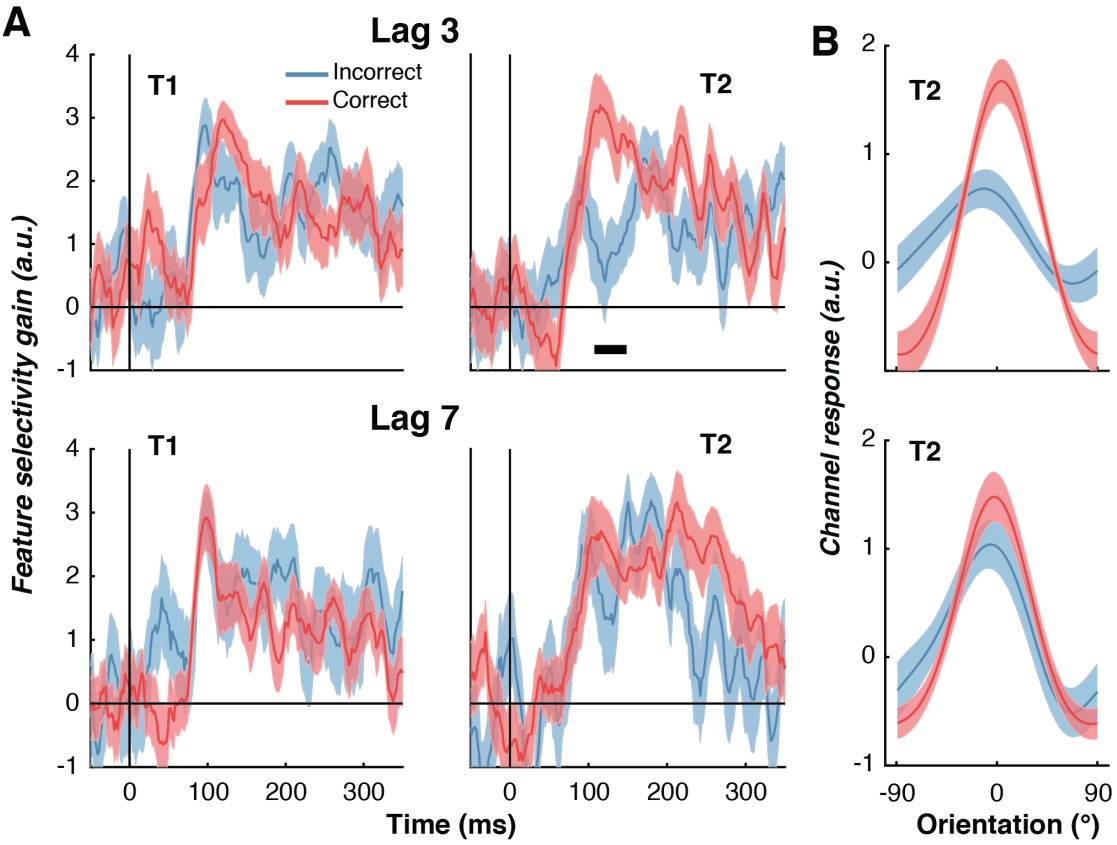
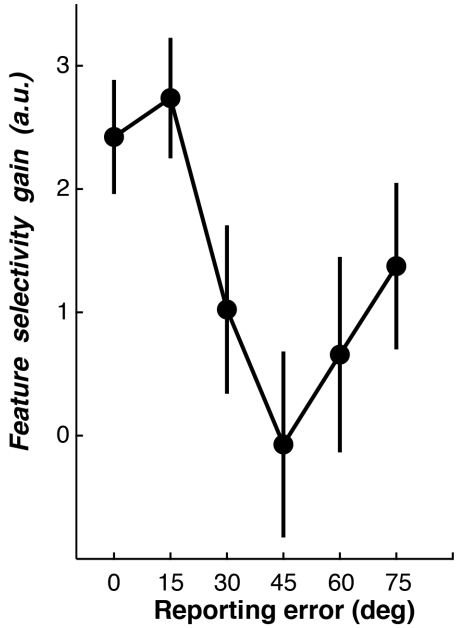


Figure 6. Neural representations of feature-selective information during the attentional blink for the first (T1) and second target (T2) based on reporting accuracy for T2 for Experiment 2. **A.** Time course of measured feature selectivity for T1 and T2, given by the gain of the fitted Gaussian parameter. Trials were scored as correct if the participant's response was within 30° of the presented T2 orientation. Only trials in which participants responded accurately to T1 were included in the analysis. The thick black horizontal line in the upper right panel indicates a period of significant difference between Incorrect and Correct trials (two-tailed sign-flipping cluster-permutation, alpha $p<0.05$, cluster alpha $p<0.05$, N permutations = 20,000). **B.** Forward encoding results were averaged across the significant time points for T2 Lag 3 shown in A (upper right panel) to reconstruct the full representation of orientation. Reliable changes in the gain of orientation representations for T2 were present at Lag 3 (upper panel) but not at Lag 7 (lower panel). There was no difference in the width for either Lag. Shading indicates ∓ 1 standard error of the mean.

To ensure we did not miss any small but consistent effects, we averaged the forward encoding results (Orientation $\times$ Time) over the early (100 to 150 ms) timepoints to increase signal-to noise-ratio and recovered the orientation tuning curve (Figure 6B).

443 Fitting Gaussians to these values confirmed that the AB was associated with a change
444 in the gain of feature selectivity for T2 at Lag 3, such that correct trials showed
445 significantly greater gain than incorrect trials ($t(22)=3.12$, $p=0.01$, $d=0.65$; Figure 6B
446 upper panel). By contrast, the width of the representation was again unaffected by the
447 AB ($t(22)=1.66$, $p=0.11$,$d=0.35$) for the same item.
448 The reduction in T2 selectivity for incorrect trials at Lag 3 was not driven by an
449 arbitrary split of trials into correct and incorrect categories. To verify this, we sorted the
450 evoked T2 forward encoding results by the amount of orientation error (in 15º error bins
451 to allow sufficient signal-to-noise ratios for fitting). There was significantly greater
452 feature selectivity when the orientation error was small, and this selectivity gradually
453 decreased with larger errors (one-way within-subjects ANOVA, $F(1,22)=2.76$, $p=0.02$,
454 $\eta_p^2=0.11$; Figure 7). These results indicate that the AB is associated with a reduction in
455 gain, but not width, of feature-selective information for the second target item (T2), and
456 that this effect occurs soon after the target appears within the RSVP stream.
457



458

459 **Figure 7.** Gain of feature-selective information for T2 items presented at Lag 3 in
460 Experiment 2, plotted as a function of reporting error. Forward encoding results were
461 averaged across early time points (100-150ms), and were binned by the absolute
462 difference between the presented and reported orientations (in 15° increments). Each
463 bin is displayed as the starting value (e.g., 0° incorporates errors from 0° to 15°).
464 Gaussians were fitted to quantify selectivity with the gain parameter shown here.
465 Feature selectivity was highest when participants reported the orientation to within 30°
466 of the presented orientation, and declined significantly with larger reporting errors. Error
467 bars indicate ∓1 standard error of the mean.
468

469 ***How do neural representations of targets and distractors contribute to the AB?***

470 We next examined the neural representations both of targets and distractors to
471 test the different predictions made by T1- versus distractor-based accounts of the AB.
472 T1-based accounts[41,42] argue that the second target deficit is caused by extended
473 processing of the first target, whereas distractor-based accounts[24,43] argue that
474 deleterious processing of the distractors, mainly between T1 and T2, causes the second
475 target to be missed. The theories thus make distinct predictions about the neural
476 representation of target and distractor items. According to T1-based accounts, target

477 representations should be enhanced relative to those of distractors, and missed T2
478 items on AB trials should be more poorly represented than correctly reported T2 items.
479 By contrast, distractor-based accounts predict that neural representations of distractor
480 items should be stronger on AB trials than on non-AB trials and weaker following T1
481 presentation.

482 As before, we averaged the forward encoding modelling representations
483 (Orientation × Time) across an early time point (100 to 150 ms), and fit Gaussians to
484 each participant's data to quantify feature selectivity (Figure 8A). For correct trials (i.e.,
485 orientation responses to T2 were within 30° of the presented orientation), the two
486 targets resulted in significantly higher feature selectivity (gain) than the immediately
487 adjacent distractors (-2,-1,+1 and +2 items) for both T1 and T2 representations (all
488 *ps*<0.04). On incorrect trials, feature selectivity for T1 was not significantly greater than
489 selectivity for the surrounding distractors ($t(22)=0.15$, $p=0.88$,$d=0.03$), even though we
490 included only trials in which T1 was correctly reported. Most interestingly, on incorrect
491 trials the representations of T2 items were significantly lower than that of the
492 immediately adjacent distractors ($t(22)=2.09$, $p=0.04$,$d=0.44$), suggesting that the
493 featural information carried by T2 was suppressed, while distractors were unaffected.
494 To directly test the distractor model of the AB, we compared distractor representations
495 before T1 with distractor representations during the AB (i.e., between T1 and T2). The
496 account predicts that distractors presented during the AB should elicit a stronger neural
497 representation as they are likely to be incorrectly selected as targets. Instead, we found
498 that distractors were represented similarly before and during the AB for both correct

trials ($t(22)=0.85$, $p=0.40$,$d=0.18$) and incorrect trials ($t(22)=1.83$, $p=0.08$,$d=0.38$).Taken together, these results suggest that for trials where participants accurately report target orientation, the neural representations of targets are boosted relative to those of distractors. By contrast, when the second target is missed, as occurs during the AB, there is a significant *suppression* of the target's featural information.

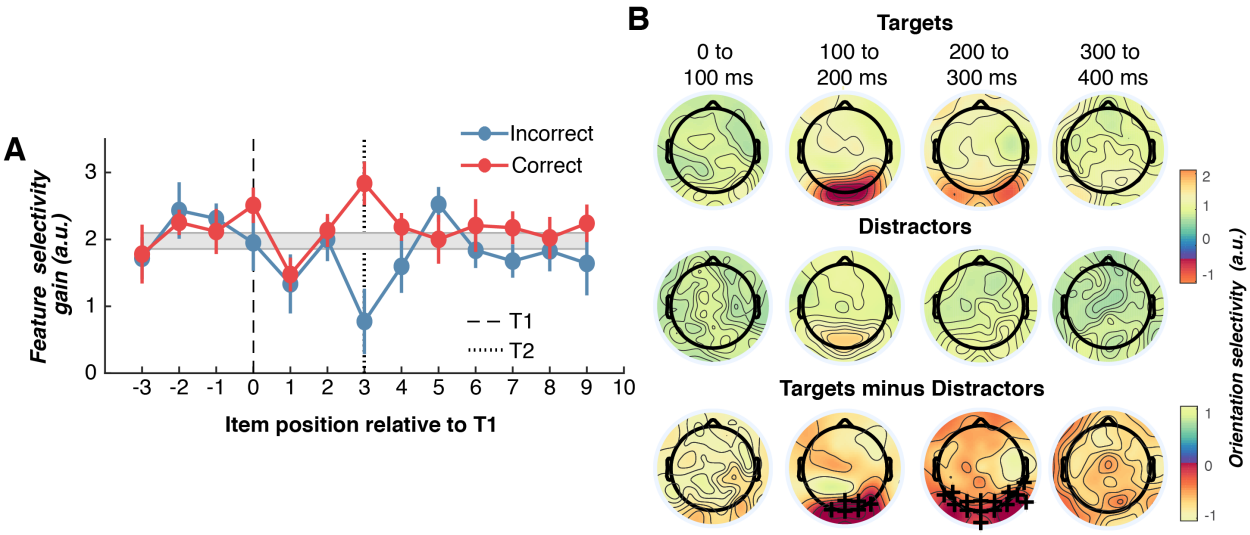


Figure 8. Feature selectivity and scalp topographies for targets and distractors in Experiment 2. **A.** Neural feature selectivity (gain of Gaussian) of target and distractor representations for Lag 3 trials. Gaussians were fit to the averaged neural representation from 100 to 150 ms. To aid comparison, the grey bar indicates the average distractor representation (∓1 standard error of mean). Note that all distractors and targets have gain values significantly above 0 arbitrary units (a.u.) indicating robust feature-selectivity. Error bars indicate ∓1 standard error of mean. **B.** Headmaps showing univariate orientation selectivity over time, plotted separately for targets and distractors. Plus symbols indicate positive cluster-permuted differences between targets and distractors (two-tailed cluster-permutation, alpha $p<0.05$, cluster alpha $p<0.025$, N permutations = 1,500).

Localisation of feature selectivity of targets and distractors

In a final step, we performed a univariate sensor-level analysis for feature selectivity[10] to find the topographies associated with target and distractor processing. To do this, we trained a simplified model of feature selectivity on each type of item (targets

and distractors) separately for each EEG sensor. Orientation information for both targets and distractors was evident most strongly over occipital and parietal areas, and target items generated significantly greater selectivity over these areas than distractors (Figure 8B). These findings suggest that while target and distractor items are processed in overlapping brain regions, targets generate significantly greater orientation-selective information than distractors.

Discussion

We developed a novel RSVP paradigm to determine the neural and behavioural bases of the limits of temporal attention. The behavioural results replicated the hallmark of the AB with response accuracy being significantly reduced when T2 was presented within 200-400 ms of T1. We discovered that target representations influenced one another, such that the reported orientation of one target was biased toward the orientation of the other. Results from Experiment 2 revealed that successfully reporting T2 depended on a boost to its neural representation relative to other items in the RSVP stream, whereas missing T2 corresponded to a suppressed neural response relative to the distractors. Notably, there was no evidence for suppression of neural representations of the distractors, suggesting the AB is primarily driven by processing competition between target items. This observation supports theories that have attributed the second-target deficit to first target processing[4,43,44], but is inconsistent with theories that attribute the AB to inadvertent processing of distractor items[24,45].

An important but unexpected result is that target reports were influenced by one another despite being separated by several hundred milliseconds and multiple distractor items. One influential theory argues that the AB is caused by temporal integration of the target with the immediate post-target distractor[25,26]. Our RSVP task found evidence for this but also showed that target representations appear to be integrated with each other even when they are separated by multiple distractor items within the stream. This finding is not explicitly predicted by any existing account of the AB. The largest bias was for Lag 1 trials, in which the two targets appear sequentially, a result that is consistent with Lag 1 switching[25,26,30]. The orientation of the immediate post-target distractor also significantly biased the perceived target orientation, whereas the distractors that appeared between the targets did not bias perceptual judgements. Taken together, our findings across two experiments suggest that the detection of a target in an RSVP sequence starts a period of local integration which involuntarily captures the next item, whether it is a target or a distractor. This is followed by a more global integration of targets, possibly within working memory[4].

Our first major aim was to determine how the AB affects target representations. The forward encoding modelling of the EEG data adds to previous results[27] by demonstrating that the gain in neural representations of Lag 3 items is significantly reduced in AB trials, compared with non-AB trials. Supporting the behavioural results, there was no effect on the width of EEG-derived feature selectivity during the AB. The neural results also go beyond the behavioural findings by showing that the gain of Lag 3 items is not only suppressed on AB trials, but boosted on non-AB trials compared with

565 those of the distractors. Taken together, these results suggest that temporal attention
566 operates in a similar manner to spatial[15,17-19] attention, but not to feature-based
567 attention[20,21], as the former has been found to affect the gain of neural responses
568 whereas the latter tends to affect the sharpness of neural tuning.

569 The second major aim of our study was to resolve the persistent debate between
570 T1- and distractor-based theories of the AB[4,24,41-45]. Behaviourally, we found scant
571 evidence that distractors (apart from the immediately subsequent distractor) influence
572 target perception. Consistent with T1-based accounts of the AB[4,24-27], there were robust
573 neural representations of distractors and no evidence that distractor representations
574 were boosted following initial target detection, as would be predicted by distractor-based
575 accounts. Furthermore, we found no evidence that post-T1 distractors were
576 suppressed, as would be predicted by T1-based inhibition accounts of the AB[4,44].
577 Instead, consistent with T1-based accounts[4,24-27], the representations of *both* targets
578 were boosted relative to those of the distractors. If the second target was missed,
579 however – as occurs during the AB – then the representation of the second target was
580 significantly suppressed relative to the distractors. Taken together, these results
581 suggest that when the first target is processed rapidly, attention is efficiently redeployed
582 to the second target, causing its representation to be boosted. By contrast, if the second
583 target appears while processing of the first target is ongoing, the visual system actively
584 suppresses the information to avoid the targets interfering with each other.

585 Suppression of the T2 representation occurred 100-150 ms after the target
586 appeared, suggesting inhibition by ongoing processing of T1. This fits well with previous

work showing that the AB is associated with a reduced late-stage response, as indicated by an ERP component associated with working memory consolidation[8,46]. Taken together with the current results, it seems that suppressing the initial T2 representation is likely to inhibit information from being passed to later cortical stages associated with working memory consolidation. As this information does not reach higher cortical stages, it results in a reduced working memory ERP component during the AB. This could also explain why the T2 representation was only initially affected (100-150 ms), as only its early appearance needs to be suppressed to stop inference with T1 processing at a higher stage. These behavioural results may be consistent with sequential working memory consolidation of targets. We found the precision of reporting T1 was unaffected by Lag, even though often during the AB only one item is reported, whereas at longer lags two items are reported. During spatial working memory tasks, where multiple items are simultaneously presented, longer lags should have a higher memory load and lead to lower precision[47]. Instead, the current results suggest that each target is consolidated into working memory before the store allows a second item to enter.

In summary, the current work adds to our understanding of the neural and behavioural basis of temporal attention. For the first time, we were able to recover a neural signature for each item within an RSVP stream, something that has not been possible with conventional approaches to EEG and fMRI data. Our novel methodology indicated that while there is co-modulation of featural information carried by each of the targets, there is no evidence for distractor suppression in this RSVP task. We also

document for the first time the existence of interactions among targets that are separated by several hundred milliseconds.

Our novel methodology provides a rich framework for exploring the neural bases of many psychological phenomena, including repetition blindness[48] and contingent attentional capture[49]. The current work was not designed to pinpoint the neural locus of the AB, combining the technique with fMRI could elucidate its cortical origins. Furthermore, feedback and feedforward processes are thought to modulate different aspects of the AB[50]. Future studies might fruitfully combine our method with invasive recordings in animals, particularly if multiple areas were recorded from simultaneously[51] to resolve these questions.

Methods

Participants

In Experiment 1, 22 participants (13 females, 9 males; median age 22 years; range 19-33 years) were recruited from a paid participant pool and reimbursed at AUD$20/hr. In Experiment 2, 23 participants (14 females, 9 males; median age 23 years; range 19-33 years old) were recruited from the same pool and reimbursed at the same rate. Each person provided written informed consent prior to participation and had normal or corrected-to-normal vision. The study was approved by The University of Queensland Human Research Ethics Committee and was in accordance with the Declaration of Helsinki.

Experimental setup

Both experiments were conducted inside dimly-illuminated rooms. The items

were displayed on a 22-inch LED monitor (resolution 1920 x 1080 pixels, refresh rate 100 Hz) using the PsychToolbox presentation software[52,53] for MATLAB. In Experiment 1, participants were seated at a distance of approximately 45 cm from the monitor. In Experiment 2, the same viewing distance was maintained using a chinrest to minimise head motion artefacts in the EEG. At a viewing distance of 45 cm, the monitor subtended 61.18º x 36.87º (one pixel = 2.4' x 2.4').

Task

A schematic of the task is shown in Figure 1. Movie 1 shows two example trials. Each trial began with a central fixation point and the RSVP stream commenced after 300 ms. The stream consisted of 20 Gabors (0.71° standard deviation, ~5° diameter, 100% contrast, centred at fixation) on a mid-grey background. On each trial, the orientations of the twenty Gabors in the stream were drawn pseudo-randomly without replacement from integer values ranging from 0-179°. Both targets and distractors were drawn from the same random distribution, meaning there was no restriction on the relationship between targets. Each item was presented for 40 ms and was separated from the next item by a blank interval of 80 ms, yielding an 8.33 Hz presentation rate. The participants' task was to reproduce the orientations of the two high-spatial frequency Gabors (targets; 2 c/°) while ignoring the items of a low-spatial frequency (distractors; 1 c/°). Between 4 and 8 distractors, varied pseudo-randomly on each trial, were presented before the first target (T1) to minimise the development of strong temporal expectations, which can reduce the AB[38,54]. The number of distractor items between T1 and T2 defined the inter-target lag (1,2,3,5,7 in Experiment 1, and 3,7 in

Experiment 2). There were 600 trials in each of the two experiments, with an equal distribution of trials across the lag conditions (120 in Experiment 1, 300 in Experiment 2), with fewer lags included in Experiment 2 to increase signal-to-noise for the regression-based EEG analysis.

Participants were asked to monitor the central RSVP stream until the presentation of the last Gabor, after which a response screen appeared (see Figure 1B). The response screen consisted of a centrally-presented black circle (10° diameter) and a yellow line. Participants rotated the line using a computer mouse to match the perceived orientation of the target and clicked to indicate their desired response. They were asked to reproduce the orientations of the two targets (T1, T2) in the order they were presented, and to respond as accurately as possible, with no time limit. After providing their responses, participants were shown a feedback screen which displayed their orientation judgements for T1 and T2, and the actual orientations of both targets (see Figure 1C). The feedback was displayed for 500 ms before the next trial began, and participants were given a self-paced rest break every 40 trials. Each experiment took between 50 and 60 minutes to complete.

EEG acquisition and pre-processing

In Experiment 2, continuous EEG data were recorded using a BioSemi Active Two system (BioSemi, Amsterdam, Netherlands). The signal was digitised at 1024 Hz sampling rate with a 24-bit A/D conversion. The 64 active scalp Ag/AgCl electrodes were arranged according to the international standard 10–20 system for electrode placement[55] using a nylon head cap. As per BioSemi system design, the common mode

675 sense and driven right leg electrodes served as the ground, and all scalp electrodes
676 were referenced to the common mode sense during recording. Pairs of flat Ag-AgCl
677 electro-oculographic electrodes were placed on the outside of both eyes, and above
678 and below the left eye, to record horizontal and vertical eye movements, respectively.
679 Offline EEG pre-processing was performed using EEGLAB in accordance with
680 best practice procedures[56,57]. The data were initially down sampled to 512 Hz and
681 subjected to a 0.5 Hz high-pass filter to remove slow baseline drifts. Electrical line noise
682 was removed using the *clean_line,* and *clean_rawdata* functions in EEGLAB[58] was used
683 to remove bad channels (identified using Artifact Subspace Reconstruction), which were
684 then interpolated from the neighbouring electrodes. Data were then re-referenced to the
685 common average before being epoched into segments for each trial (-0.5 s to 3.0 s
686 relative to the first Gabor in the RSVP). Systematic artefacts from eye blinks,
687 movements and muscle activity were identified using semi-automated procedures in the
688 SASICA toolbox[59] and regressed out of the signal. The data were then baseline
689 corrected to the mean average EEG activity from 500 to 0 ms before the first Gabor in
690 the trial.

691 **Behavioural analysis**

692 To determine how the AB affected participants' perception of targets, for each
693 trial we found the difference between the actual target orientation and the reported
694 orientation (i.e., the *orientation error*) for T1 and T2. This approach is analogous to one
695 employed in previous work that examined whether the AB is associated with discrete or
696 graded awareness of T2[27]. The continuous nature of the orientation responses given by

participants on each trial raises the challenge of distinguishing "correct" and "incorrect" trials. For Experiment 2, we scored trials as correct when the orientation error was less than 30° from the presented orientation; trials were scored as incorrect when the orientation error was greater than 30°. As shown in Supplementary Figure 5, this approach to scoring yielded a classic blink effect, suggesting the tasks captures the important behavioural features of the widely-reported AB phenomenon. For each lag condition, we found the proportion of responses (in 15° bins) between -90° and +90° for the orientation errors (see Figure 2A and 4A) and fit Gaussian functions with a constant offset (Equation 1) using non-linear least square regression to quantify these results for each participant (Figure 2B and 4B):

$$G(x) = A\, exp(-\frac{(x-\mu)^2}{2\sigma^2}) + C \quad (1)$$

Where *A* is the gain, reflecting the number of responses around the reported orientation, μ is the orientation on which the function is centred (in degrees), σ is the standard deviation (degrees), which provides an index of the precision of participants' responses, and *C* is a constant used to account for changes in the guessing rate. Using different bin sizes yields the same pattern of results suggesting this procedure did not bias the results (Supplementary Figure 3).

We used a regression-based approach (see Figure 2D and 4C) to determine how targets and distractors within each RSVP stream influenced behavioural responses[60]. To do this, we aligned the orientations of both distractor and target items from 4 items prior to the appearance of the target through to 9 items after the appearance of the target to construct a regression matrix of the presented orientations. The regression

719 matrix was converted to complex numbers (to account for circularity of orientations)
720 using Equation 2.

721 $$C_1 = \exp(1i\ C) \quad (2)$$

722 Where C is the regression matrix (in radians) and 1i is an imaginary unit. Standard
723 linear regression was used to determine how the orientations of the items affected the
724 reported orientation using Equation 3.

725 $$W = (C_1\ C_1^T)^{-1}\ C_1^T\ R \quad (3)$$

726 Where R is the reported orientation (in radians). This was done separately for T1 and
727 T2 reports, with a higher regression weight indicating the item was more influential in
728 determining the reported orientation.

729 To determine whether the finding that the orientations of T1 and T2 influenced
730 the reported orientation was due to participants integrating the other target or the
731 surrounding distractors (Figure 3), we found the difference in orientation between the
732 target of interest and the other item (either target or distractor) and the orientation error
733 for each trial. This showed an orientation-tuned effect characteristic of integration. To
734 quantitatively determine the magnitude of this effect, we fit first-derivative Gaussian
735 functions (D1; Equation 4) to these responses[35,36].

736 $$\mathrm{D1}(x) = A \times \frac{1}{\sigma} \times x - \mu \times exp(-\frac{x-\mu^2}{2\sigma^2}) \quad (4)$$

737 Where *A* is the gain, μ is the orientation on which the function is centred (in degrees),
738 and σ is the standard deviation (degrees).

739 **Forward encoding modelling**

740 Forward encoding modelling was used to recover orientation-selective responses

741 from the pattern of EEG activity for both target and distractor items in the RSVP stream.
742 This technique has been used previously to reconstruct colour[16], spatial[15] and
743 orientation[19] selectivity from timeseries data acquired through functional magnetic
744 resonance imaging (fMRI). More recently, the same approach has been used to encode
745 orientation[9-13] and spatial[14] information contained within MEG and EEG data, which
746 have better temporal resolution than fMRI.
747 We used the orientations of the epoched data segments to construct a
748 regression matrix with 9 regression coefficients, one for each of the orientations (Figure
749 9A). This regression matrix was convolved with a tuned set of nine basis functions (half
750 cosine functions raised to the eighth power[9,10,19], Equation 5) centred from 0° to 160° in
751 20° steps.

752 $$F(x) = \cos(x - \mu)^8 \quad (5)$$

753 Where μ is the orientation that the channel is centred, and x are orientations from 0º to
754 180º in 1º steps.
755

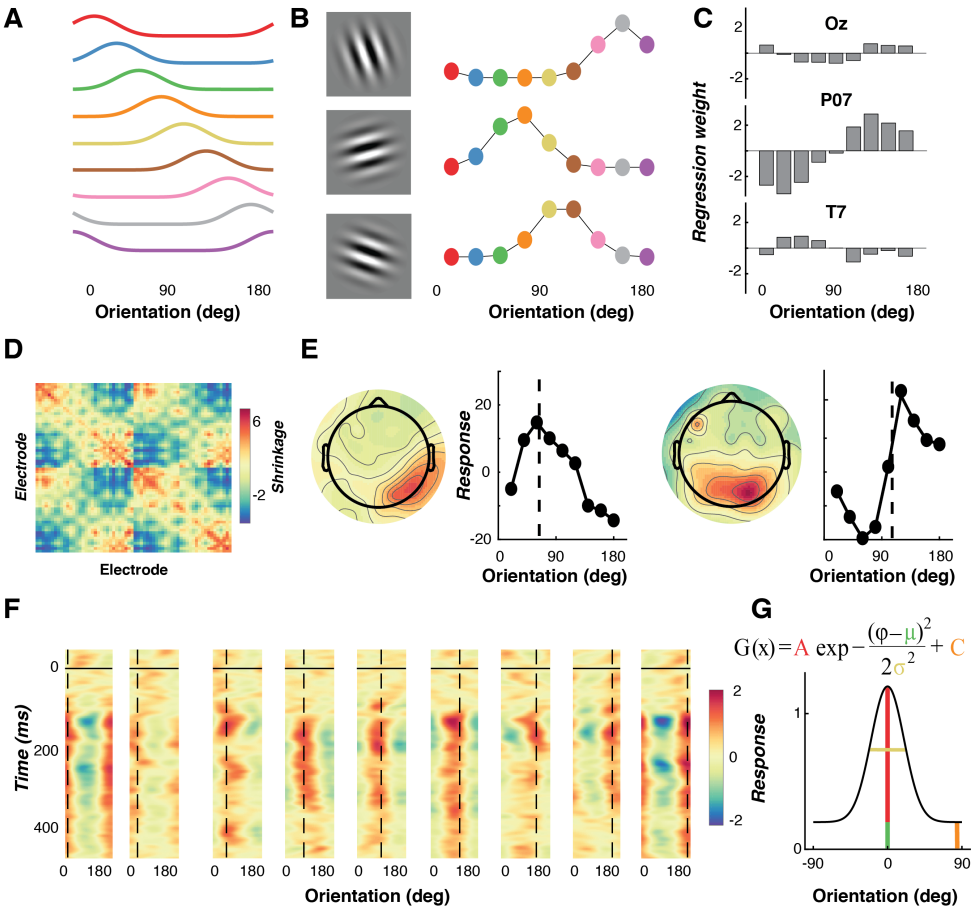


Figure 9. Schematic illustrating the forward encoding procedure used to estimate feature-selectivity for orientation in Experiment 2. **A.** A basis set of the nine channels used to model feature (orientation) selectivity. **B.** The basis set was used to find the expected response (regression coefficients) for each different RSVP item in every trial, for each EEG electrode (three electrodes are shown here for a single example participant). Three trials are shown for the corresponding gratings. **C**. Ordinary least squares regression was used to find regression weights for the orientation channels across trials for each EEG electrode (three electrodes are shown here for a single example participant). **D.** Shrinkage matrix[11,13,61] that the weights were divided by to perform regularization, to account for correlated activity between electrodes. **E.** The regression weights were applied to predict the presented orientation. Neural activity (headmaps) from two trials, with the channel responses for those trials. Dotted lines indicate the presented orientations. **F.** Applying this procedure to each time point gives the time-course of feature-(orientation) selectivity (for one participant). Trials have been binned in 20º intervals, with the dotted lines representing the presented orientation in those trials. On the y-axis, 0 ms represents the onset of the item within the RSVP stream. Feature selectivity emerged around 75 ms after stimulus presentation. **G.** Modified Gaussian functions (equation) were used to quantify the tuning. The colours of the free parameters in the equation correspond to the relevant components of the tuning curve below.

This tuned regression matrix was used to measure orientation information either across trials or in epoched segments. This was done by solving the linear equation (6):

$$B_1 = WC_1 \quad (6)$$

Where B_1 (64 sensors $\times$ N training trials) is the electrode data for the training set, C_1 (9 channels $\times$ N training trials) is the tuned channel response across the training trials, and W is the weight matrix for the sensors to be estimated (64 sensors x 9 channels). Following methods recently introduced for M/EEG analysis, we separately estimated the weights associated with each channel individually[13,61]. W was estimated using least square regression to solve equation (7):

$$W = (C_1 C_1^T)^{-1} C_1^T B_1 \quad (7)$$

Following this previous work[11,13,61], we removed the correlations between sensors, as these add noise to the linear equation. To do this, we first estimated the noise correlation between electrodes and removed this component through regularisation[62,63] by dividing the weights by the shrinkage matrix. The channel response in the test set C_2 (9 channels $\times$ N test trials) was estimated using the weights in (7) and applied to activity in B_2 (64 sensors $\times$ N test trials), as per Equation 8:

$$C_2 = (W W^T) W^T B_2 \quad (8)$$

To avoid overfitting, we used cross validation (10-fold in the initial whole-trial analysis, and 20-fold when the item presentations were stacked), where X-1 of epochs were used to train the model, and this was then tested on the remaining (X) epoch. This process was repeated until all epochs had served as both test and training trials. We also repeated this procedure for each point in the epoch to determine time-resolved feature-

799 selectivity. To re-align the trials with the exact presented orientation, we reconstructed
800 the item representation[15] by multiplying the channel weights (9 channels × time × trial)
801 against the basis set (180 orientations × 9 channels). This resulted in a 180 (-89° to 90°)
802 Orientations × Trial × Time reconstruction. In order to average across trials, the
803 orientation dimension was shifted so that 0º corresponded to the presented orientation
804 in each trial.

805 For the initial encoding analysis (Figure 5), to determine whether feature
806 selectivity could be recovered for each RSVP item we used 20 encoding models (one
807 for each item position in the stream) with 600 trials. We trained and tested each model
808 across the entire 2250 ms of the trial to determine when feature selectivity emerged for
809 that RSVP item. This analysis verified that each RSVP item could be encoded
810 independently. We aligned all RSVP items across trials (N = 12,000; 600 trials by 20
811 items) and used a fixed encoding model for training and testing (Figures 6-8)[64,65]. This
812 meant we trained and tested all encoding models across all items (both targets and
813 distractors) regardless of trial type[12,13].

814 Aligned item reconstructions were then averaged over the relevant condition
815 (Lag, Accuracy or item position) and smoothed using a Gaussian with a temporal kernel
816 of 6 ms[10,12] to quantify feature selectivity. The Gaussian functions were fit, using least
817 square regression, to quantify different parameters of feature selectivity across time
818 points, as per Equation 1, where A is the gain representing the amount of feature
819 selective activity, μ is the orientation on which the function is centred (in degrees), σ is
820 the width (degrees), and C is a constant used to account for non-feature selective

821 baseline shifts.

822 **Univariate orientation selectivity analysis**

823 We used a univariate selectivity analysis[10] to determine the topography
824 associated with orientation-selective activity for targets and distractors (Figure 8B). Data
825 were epoched in the same manner as in the forward encoding model where EEG
826 activity was aligned with each stream item. We separated these epochs into target and
827 distractor presentations to determine whether these two types of stimulus were
828 processed differently. All target presentations were used in training (1200 in total; 600
829 trials with two targets in each), together with a pseudo-random selection of the same
830 number of distractor items. To determine the topography, we used a general linear
831 model to estimate orientation selectivity for each sensor from the sine and cosine of the
832 presentation orientation, and a constant regressor in each presentation. From the
833 weights of the two orientation coefficients we calculated selectivity using Equation 9:

834 $$A = \sqrt{B_1 cos^2 + B_2 sin^2} \quad (9)$$

835 *A* was derived through permutation testing in which the design matrix was shuffled (N =
836 1000) and weights calculated. The non-permuted weights were ranked and compared
837 with the permutation distribution, thus enabling calculation of the z-scored difference. To
838 calculate group-level effects, cluster-based sign-flipping permutation testing (N=1500)
839 across electrodes and time was implemented in Fieldtrip[66] to determine whether the
840 topographies differed between conditions.

841 **Statistics**

842 All statistical tests were two-sided, and Bonferroni adjustments were used to

correct for multiple comparisons where noted. Non-parametric sign permutation tests[67,68] were used to determine differences in the time courses of feature selectivity (Figures 5 and 6) between conditions. The sign of the data was randomly flipped (N=20,000), with equal probability, to create a null distribution. Cluster-based permutation testing was used to correct for multiple comparisons over the time series, with a cluster-form threshold of $p < .05$ and significance threshold of $p < .05$.

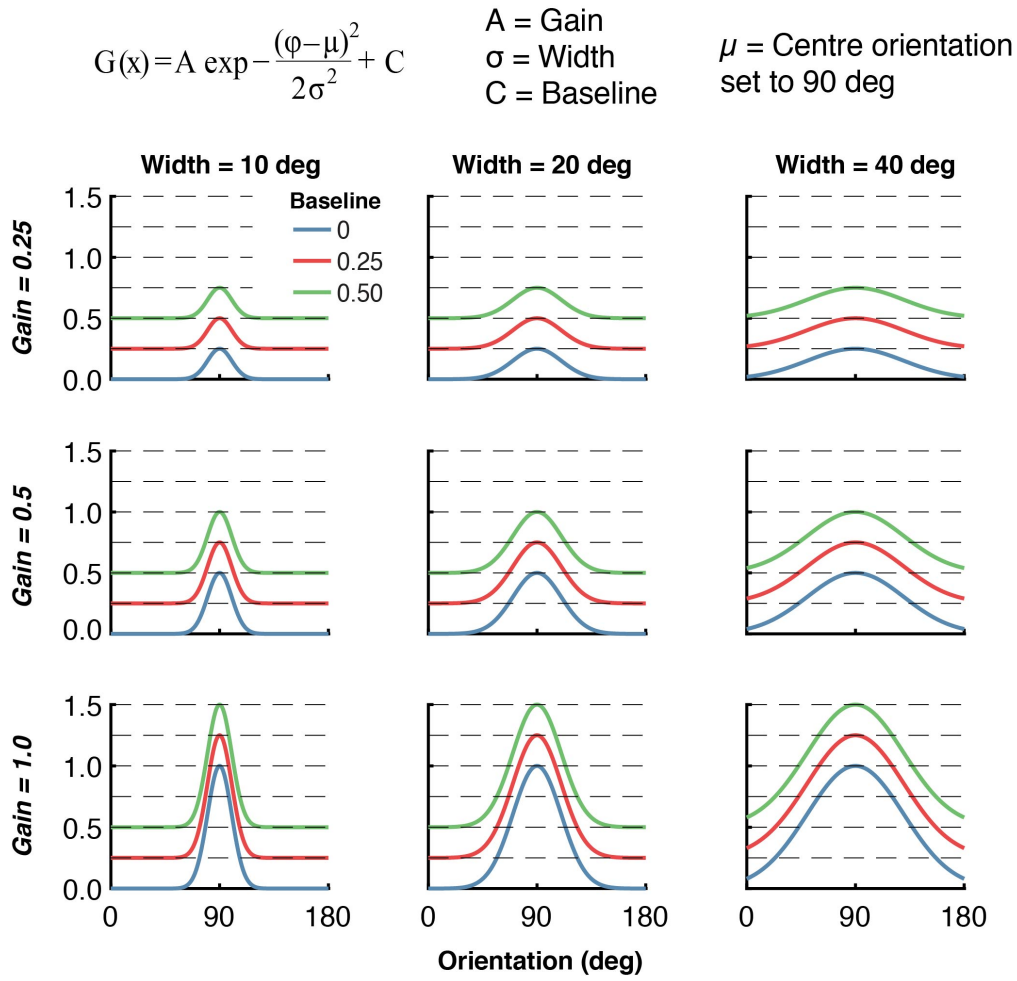
Data availability

The EEG and behavioural data for both experiments are available at: https://osf.io/f9g6h. The code is available at: https://github.com/MatthewFTang/AttentionalBlinkForwardEncoding.

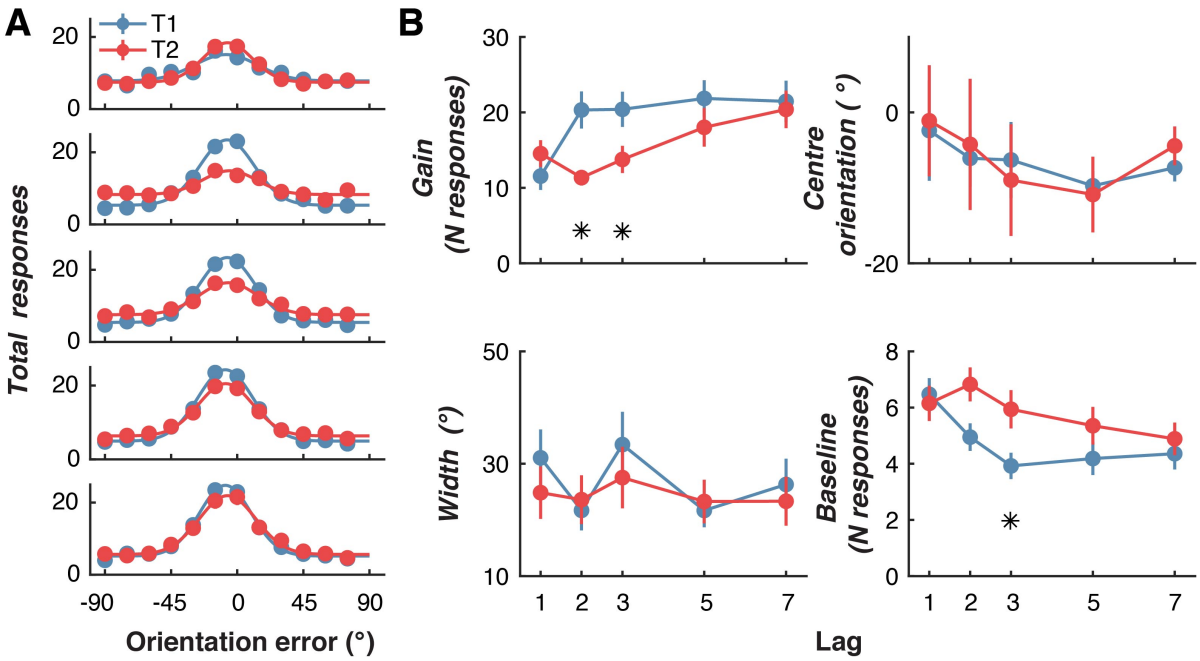
Acknowledgements

We would like to thank Brad Wyble for comments on a pre-print of this manuscript. This work was supported by the Australian Research Council (ARC) Centre of Excellence for Integrative Brain Function (CE140100007) to JBM and EA. MFT was supported by the NVIDIA corporation who donated a TITAN V GPU. EA was supported by an ARC Discovery Project (DP170100908). JTE was supported by a Discovery Grant from the Natural Sciences and Engineering Research Council (Canada). TV was supported by an ARC Discovery Project (DP120102313). JBM was supported by an ARC Australian Laureate Fellowship (FL110100103) and by the Canadian Institute for Advanced Research (CIFAR).

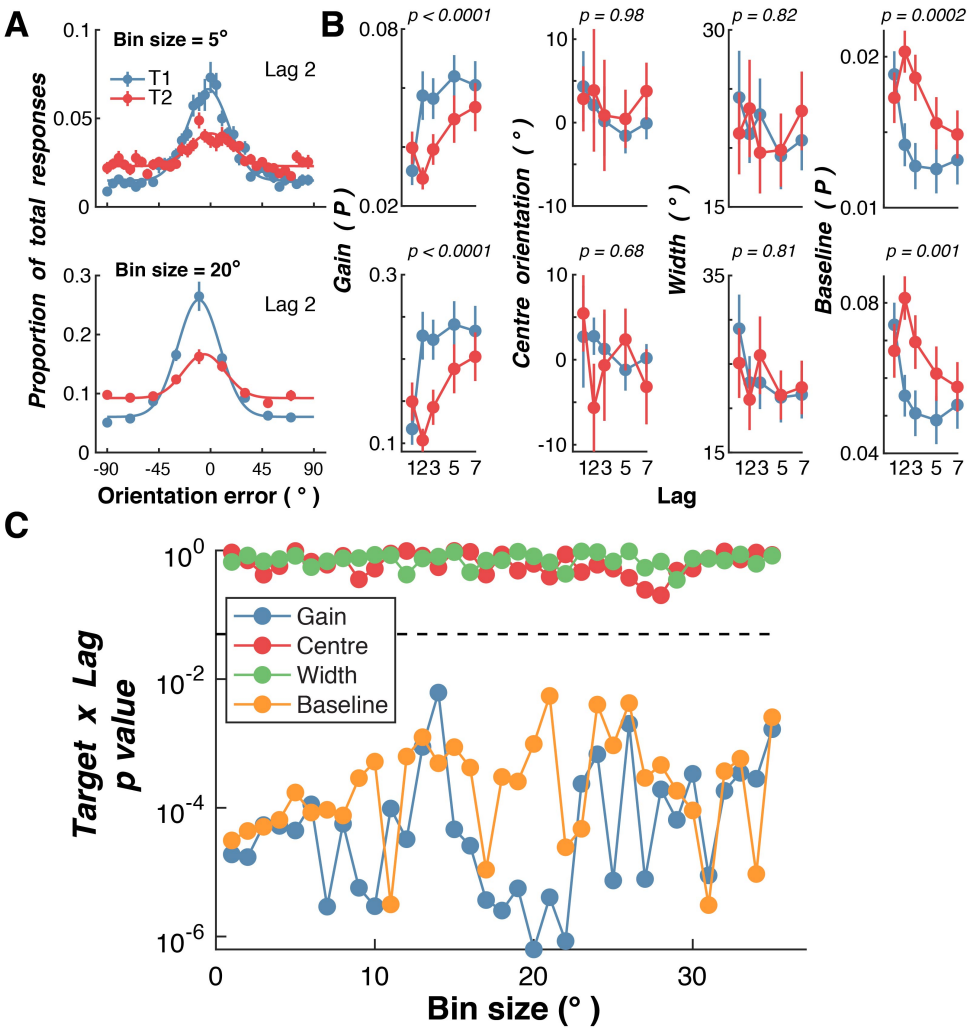
Supplementary material



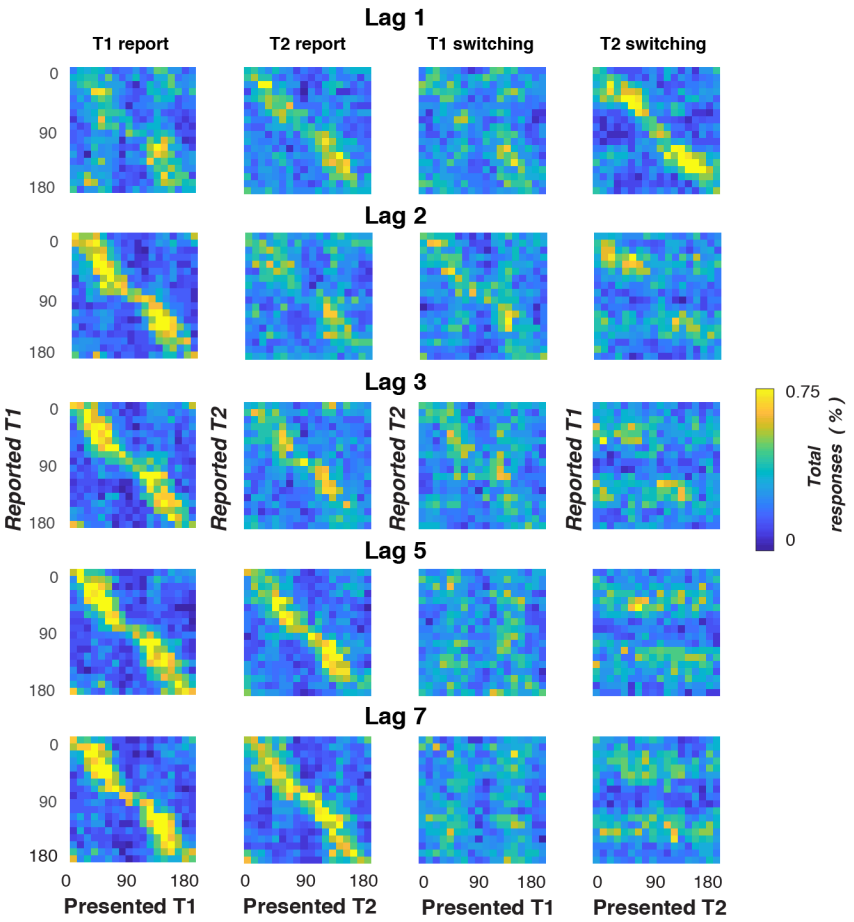
Supplementary Figure 1. Examples of the four-parameter Gaussians used to quantify behavioural orientation errors and forward encoding representations for orientation from EEG activity. The figure shows the effect of different parameter values on the shape of the resulting function. Each row has a different gain value, and each column has a different width parameter. Within each panel, the baseline value changes. The width parameter shows the precision (of either the behavioural responses or neural representations). The baseline parameter captures non-selective responses that are unrelated to the target. For the behavioural analysis, this reflects random guessing which would be distributed equally across all orientations; for the EEG analysis, it reflects overall, non-feature selective activity from the orientation encoding. For all panels, the centre orientation of the Gaussian is set to 90°. The figure highlights the independence of the parameters of the Gaussians. For instance, looking at the panels across a given row (where width varies, but gain is fixed) reveals that curves with the same baseline value have peaks at the same height. Inspecting any one panel shows that the baseline and gain parameters are independent, with the differences between the peaks of the curves being equal regardless of the baseline.



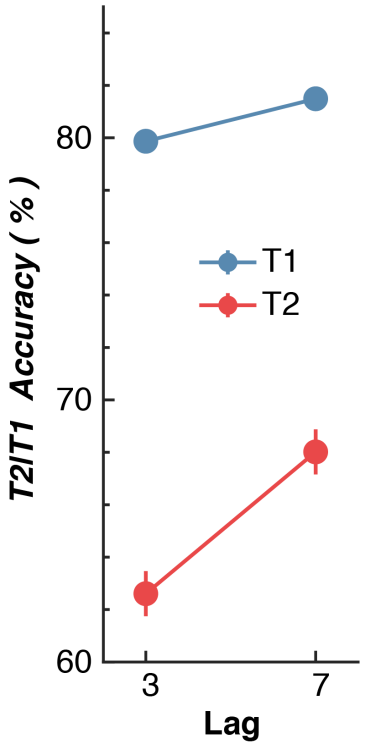
Supplementary Figure 2. Re-analysis of behavioural results in Experiment 1. In the original analysis (Figure 2), the responses were normalized to 1 for each participant to show the proportion of total responses in each orientation error bin. To confirm that this normalization did not bias the results, we re-ran the analysis but used the total number of responses (120 trials for each condition). **A.** The distribution of response errors (difference between presented and reported orientation) across participants for T1 and T2 for each Lag condition. Lines show fitted Gaussian functions. **B.** Quantified behavioural responses for the four parameters of the fitted Gaussian function for each participant. We used 2 (Target; T1,T2) x 5 (Lag; 1,2,3,5,7) within-subject ANOVAs to quantify the effect of the AB on the parameters of the fitted Gaussians. The gain parameter was affected by the factors of Target ($F(1,21)=8.97$, $p=0.007$, $\eta_p^2=0.30$) and Lag ($F(4,84)=11.78$, $p<.0001$, $\eta_p^2=0.36$), and there was a significant interaction between these factors ($F(4,84)=7.29$, $p<.0001$, $\eta_p^2=0.26$). By contrast, and consistent with the original analysis, for the width parameter there were no significant main effects of Target ($F(1,21)=0.54$, $p=0.47$, $\eta_p^2=0.02$) or Lag ($F(4,84)=1.08$, $p=0.37$, $\eta_p^2=0.05$), and no interaction($F(4,84)=0.60$, $p=0.66$, $\eta_p^2=0.03$). The baseline parameter, which reflects guessing of random orientations, was significantly affected by Target ($F(1,21)=8.72$, $p=0.008$, $\eta_p^2=0.29$) and Lag ($F(4,84)=3.54$, $p=0.01$, $\eta_p^2=0.14$). There was also a significant interaction between these factors ($F(4,84)=3.04$, $p=0.02$, $\eta_p^2=0.13$). Taken together, the results replicate those reported in the main analysis, and confirm that the process of normalization did not bias the outcomes. Asterisks indicate Bonferroni-corrected differences at $p < .05$. Error bars indicate ∓1 standard error of mean.



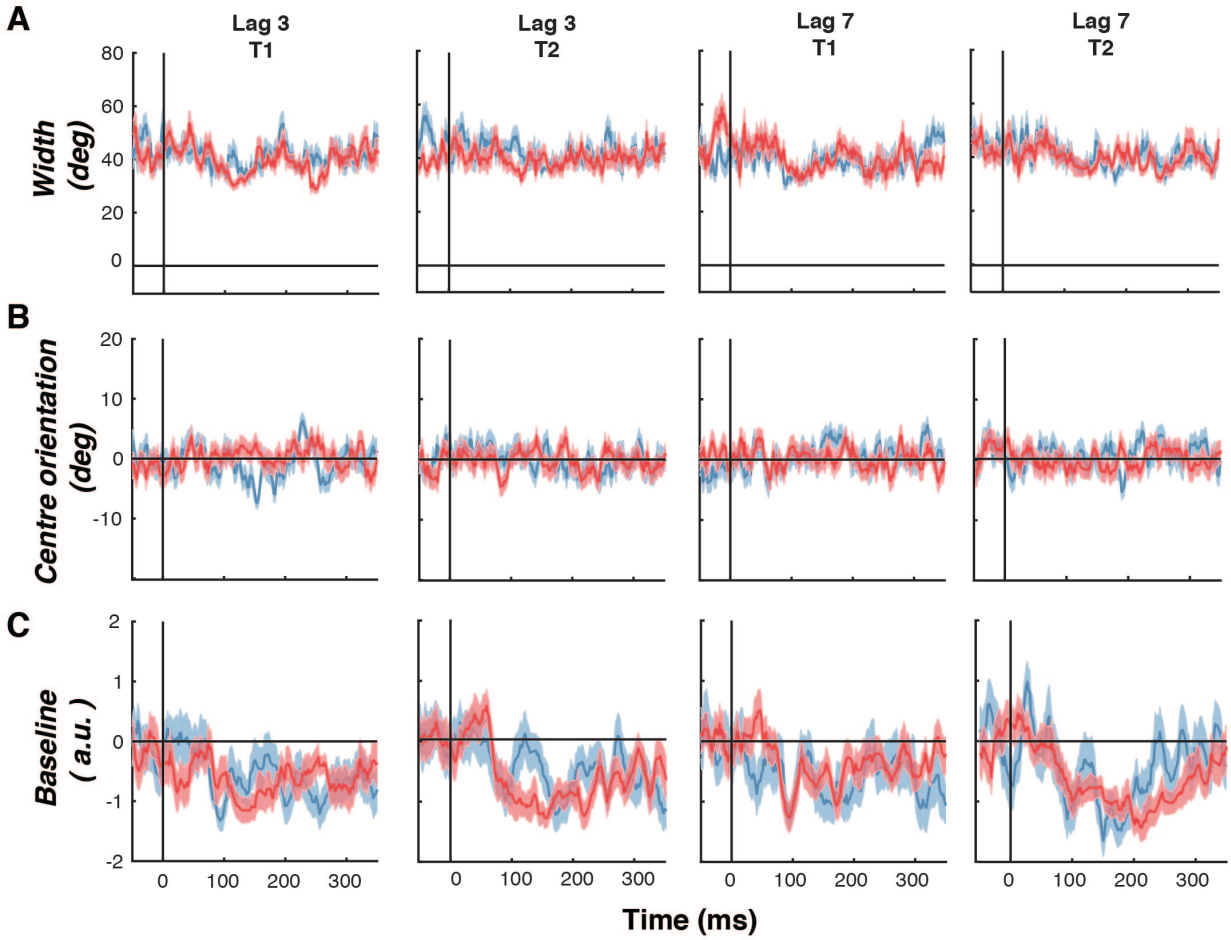
Supplementary Figure 3. Re-analysis of behavioural results from Experiment 1, with different bin sizes of orientation errors to generate the orientation error histogram. In the original analysis, a 15° bin size was used to group responses. **A.** Examples of Lag 2 histograms for two bin sizes (5° and 20°) for responses across participants. For the 5° bin size, orientation errors from -90° to -86° would be grouped together. By contrast, for a 20° bin, orientation errors from -90° to -71° would be grouped together. **B.** For each participant, Gaussians were fit to the resulting response function for each Lag and Target (T1 and T2). Here the fits are shown for 5° (top row) and 20° (bottom row) for the four parameters of the Gaussian. The p value is for the interaction term from the within-subjects ANOVA used in the original analysis, with factors of Lag (1,2,3,5,7) x Target (T1, T2). In the original analysis, as in the classic AB, a significant Lag x Target interaction shows that T2 accuracy is impaired at early Lags, whereas T1 accuracy is unaffected by Lag. **C.** P-values for the interaction term (Lag x Target) across a wide range of bin sizes for the four Gaussian parameters. The dotted line indicates p = 0.05. Note that the p values are displayed on a log axis. As in the original analysis, across this wide range of bin sizes, there was a clear AB effect on the gain and baseline parameters of the Gaussian, but no such effect on the width or centre orientation parameters.



Supplementary Figure 4. Heatmaps of reported versus presented orientations across all participants in Experiment 1. These are analogous to confusion matrices for a continuous report task. Warmer colours indicate a greater proportion of responses. Note that participants were asked to report the targets in their presented order. Each row is a lag and each column shows a different comparison. Panels in the leftmost column show presented-T1 orientation against reported-T1 orientation. For Lags 2-7, there is a strong correspondence between presented and reported orientations, confirming that participants accurately reported T1 targets. The next column to the right shows the outcome of the same analysis, but for T2 targets. For these items, there was a strong correspondence between presented and reported orientations at Lags 1, 5 and 7, which decreases (i.e., more random) for items at Lags 2 and 3. The next column shows T1 switching, where presented-T1 orientation is plotted against reported-T2 orientation. The rightmost column shows T2 switching, where presented-T2 orientation is plotted against reported-T2 orientation. Clear switching is evident only at Lag 1, where the orientation of the item presented at T2 is reported as the orientation of T1.



Supplementary Figure 5. Behavioural accuracy for the RSVP task in Experiment 2. Each response was scored as correct if the participant responded within ±30° of the presented orientation. Trials were only included if participants responded correctly (within ±30°) to T1. Following conventional AB analysis procedures, T2 accuracy was scored using only those trials in which the T1 response was correct. Error bars show within-subject standard error.



Supplementary Figure 6. Plots of three fitted parameters for the neural representations of feature-selective information over time for T1 and T2 items in Experiment 2. **A.** Time course of measured width of feature selectivity for T1 and T2 items, given by the width (standard deviation) of the fitted Gaussian. Trials were scored as correct if the participant's response was within 30° of the presented orientation. Only trials in which participants responded correctly to T1 were included in the analysis. **B.** Same as in panel **A** but for the centre orientation parameter. **C.** Same as in panel **A** but for the baseline parameter. For all panels, there were no significant differences between conditions (two-tailed cluster-permutation, alpha $p < .05$, cluster alpha $p < .05$, N permutations = 20,000). Shading indicates ∓ 1 standard error of mean.

References

1. McLeod, P. Parallel processing and the psychological refractory period. *Acta Psychologica* **41,** 381–396 (1977).
2. Welford, A. T. The 'Psychological refractory period' and the timing of high-speed performance—A review and a theory. *British Journal of Psychology. General Section* **43,** 2–19 (1952).
3. Shapiro, K. L., Raymond, J. E. & Arnell, K. M. The attentional blink. *Trends Cogn Sci* **1,** 291–296 (1997).
4. Raymond, J. E., Shapiro, K. L. & Arnell, K. M. Temporary suppression of visual processing in an RSVP task: an attentional blink? . *J Exp Psych: HPP* **18,** 849–860 (1992).
5. Broadbent, D. E. & Broadbent, M. H. P. From detection to identification: Response to multiple targets in rapid serial visual presentation. *Percept Psychophys* **42,** 105–113 (1987).
6. Williams, M. A., Visser, T. A. W., Cunnington, R. & Mattingley, J. B. Attenuation of Neural Responses in Primary Visual Cortex during the Attentional Blink. *J Neurosci* **28,** 9890–9894 (2008).
7. Marois, R., Yi, D.-J. & Chun, M. M. The Neural Fate of Consciously Perceived and Missed Events in the Attentional Blink. *Neuron* **41,** 465–472 (2004).
8. Vogel, E. K., Luck, S. J. & Shapiro, K. L. Electrophysiological evidence for a postperceptual locus of suppression during the attentional blink. *J Exp Psych: HPP* **24,** 1656–1674 (1998).
9. Garcia, J. O., Srinivasan, R. & Serences, J. T. Near-real-time feature-selective modulations in human cortex. *Current Biology* **23,** 515–522 (2013).
10. Myers, N. E. *et al.* Testing sensory evidence against mnemonic templates. *eLife Sciences* **4,** e09000 (2015).
11. Wolff, M. J., Jochim, J., Akyürek, E. G. & Stokes, M. G. Dynamic hidden states underlying working-memory-guided behavior. **20,** 864–871 (2017).
12. Tang, M. F., Smout, C. A., Arabzadeh, E. & Mattingley, J. B. Prediction error and repetition suppression have distinct effects on neural representations of visual information. *eLife Sciences* **7,** e33123 (2018).
13. Smout, C. A., Tang, M. F., Garrido, M. I. & Mattingley, J. B. Attention promotes the neural encoding of prediction errors. *PLoS Biol* **17,** e2006812 (2019).
14. Foster, J. J., Sutterer, D. W., Serences, J. T., Vogel, E. K. & Awh, E. The topography of alpha-band activity tracks the content of spatial working memory. *J Neurophysiol* **115,** 168–177 (2016).
15. Sprague, T. C. & Serences, J. T. Attention modulates spatial priority maps in the human occipital, parietal and frontal cortices. *Nat Neurosci* **16,** 1879–1887 (2013).
16. Brouwer, G. J. & Heeger, D. J. Decoding and reconstructing color from responses in human visual cortex. *J Neurosci* **29,** 13992–14003 (2009).
17. McAdams, C. J. & Maunsell, J. H. R. Effects of Attention on Orientation-Tuning Functions of Single Neurons in Macaque Cortical Area V4. *J Neurosci* **19,** 431–441 (1999).

18. Maunsell, J. Neuronal Mechanisms of Visual Attention. *Annu. Rev. Vis. Sci.* **1,** 373–391 (2015).
19. Ester, E. F., Sutterer, D. W., Serences, J. T. & Awh, E. Feature-Selective Attentional Modulations in Human Frontoparietal Cortex. *J Neurosci* **36,** 8188–8199 (2016).
20. Treue, S. & Trujillo, J. Feature-based attention influences motion processing gain in macaque visual cortex. *Nature* (1999). doi:10.1038/21176
21. Maunsell, J. H. R. & Treue, S. Feature-based attention in visual cortex. *Trends Neurosci* **29,** 317–322 (2006).
22. Treue, S. Neural correlates of attention in primate visual cortex. *Trends Neurosci* **62,** 117 (2001).
23. Maloney, R. T., Jayakumar, J., Levichkina, E. V., Pigarev, I. N. & Vidyasagar, T. R. Information processing bottlenecks in macaque posterior parietal cortex: an attentional blink? *Exp Brain Res* **228,** 365–376 (2013).
24. Di Lollo, V., Kawahara, J.-I., Shahab Ghorashi, S. M. & Enns, J. T. The attentional blink: resource depletion or temporary loss of control? *Psychol Res* **69,** 191–200 (2005).
25. Akyürek, E. G. & Hommel, B. Target integration and the attentional blink. *Acta Psychologica* **119,** 305–314 (2005).
26. Akyürek, E. G. *et al.* Temporal target integration underlies performance at lag 1 in the attentional blink. *J Exp Psych: HPP* **38,** 1448–1464 (2012).
27. Asplund, C. L., Fougnie, D., Zughni, S., Martin, J. W. & Marois, R. The attentional blink reveals the probabilistic nature of discrete conscious perception. *Psych Sci* **25,** 824–831 (2014).
28. Dehaene, S. & Naccache, L. Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. *Cognition* (2001). doi:10.1016/S0010-0277(00)00123-2
29. Nieuwenhuis, S. & de Kleijn, R. Consciousness of targets during the attentional blink: a gradual or all-or-none dimension? *Atten Percept Psychophys* **73,** 364–373 (2011).
30. Visser, T. A. W. Expectancy-Based Modulations of Lag-1 Sparing and Extended Sparing During the Attentional Blink. *J Exp Psych: HPP* 1–17 (2015). doi:10.1037/a0038754
31. Alais, D., Leung, J. & Van der Burg, E. Linear Summation of Repulsive and Attractive Serial Dependencies: Orientation and Motion Dependencies Sum in Motion Perception. *J Neurosci* **37,** 4381–4390 (2017).
32. Fischer, J. & Whitney, D. Serial dependence in visual perception. *Nat Neurosci* **17,** 738–743 (2014).
33. Fritsche, M., Mostert, P. & de Lange, F. P. Opposite Effects of Recent History on Perception and Decision. *Current Biology* **27,** 590–595 (2017).
34. Dickinson, J. E., Almeida, R. A., Bell, J. & Badcock, D. R. Global shape aftereffects have a local substrate: A tilt aftereffect field. *J Vis* **10,** 5–5 (2010).
35. Dickinson, J. E., Morgan, S. K., Tang, M. F. & Badcock, D. R. Separate banks of information channels encode size and aspect ratio. *J Vis* **17,** 27 (2017).

36. Tang, M. F., Dickinson, J. E., Visser, T. A. W. & Badcock, D. R. The broad orientation dependence of the motion streak aftereffect reveals interactions between form and motion neurons. *J Vis* **15,** 4 (2015).
37. Visser, T. A. W., Bischof, W. F. & Di Lollo, V. Rapid serial visual distraction: task-irrelevant items can produce an attentional blink. *Percept Psychophys* **66,** 1418–1432 (2004).
38. Visser, T. A. W., Tang, M. F., Badcock, D. R. & Enns, J. T. Temporal cues and the attentional blink: a further examination of the role of expectancy in sequential object perception. *Atten Percept Psychophys* **76,** 2212–2220 (2014).
39. Visser, T. A. W. T1 difficulty and the attentional blink: Expectancy versus backward masking. *The Quarterly Journal of Experimental Psychology* **60,** 936–951 (2007).
40. Brehaut, J. C., Enns, J. T. & Di Lollo, V. Visual masking plays two roles in the attentional blink. *Percept Psychophys* **61,** 1436–1448 (1999).
41. Chun, M. M. & Potter, M. C. A two-stage model for multiple target detection in rapid serial visual presentation. *J Exp Psych: HPP* **21,** 109–127 (1995).
42. Duncan, J., Ward, R. & Shapiro, K. Direct measurement of attentional dwell time in human vision. *Nature* **369,** 313–315 (1994).
43. Wyble, B., Bowman, H. & Nieuwenstein, M. The attentional blink provides episodic distinctiveness: Sparing at a cost. *J Exp Psych: HPP* **35,** 787–807 (2009).
44. Raymond, J. E., Shapiro, K. L. & Arnell, K. M. Similarity determines the attentional blink. *J Exp Psych: HPP* **21,** 653–662 (1995).
45. Lagroix, H. E. P., Spalek, T. M., Wyble, B., Jannati, A. & Di Lollo, V. The root cause of the attentional blink: First-target processing or disruption of input control? *Atten Percept Psychophys* **74,** 1606–1622 (2012).
46. Shapiro, K., Driver, J. & Ward, R. Priming from the attentional blink: A failure to extract visual tokens but not visual types. *Psych Sci* (1997).
47. Bays, P. M., Catalao, R. F. G. & Husain, M. The precision of visual working memory is set by allocation of a shared resource. *J Vis* **9,** 7–7 (2009).
48. Johnston, J. C., Hochhaus, L. & Ruthruff, E. Repetition blindness has a perceptual locus: evidence from online processing of targets in RSVP streams. *J Exp Psych: HPP* **28,** 477–489 (2002).
49. Folk, C. L., Remington, R. W. & Johnston, J. C. Involuntary covert orienting is contingent on attentional control settings. *J Exp Psych: HPP* **18,** 1030–1030 (1992).
50. Hanslmayr, S., Gross, J., Klimesch, W. & Shapiro, K. L. The role of alpha oscillations in temporal attention. *Brain Res Rev* **67,** 331–343 (2011).
51. Jun, J. J. *et al.* Fully integrated silicon probes for high-density recording of neural activity. *Nature* **551,** 232–236 (2017).
52. Pelli, D. G. The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spat Vis* **10,** 437–442 (1997).
53. Brainard, D. H. The Psychophysics Toolbox. *Spat Vis* **10,** 433–436 (1997).

54. Tang, M. F., Badcock, D. R. & Visser, T. A. W. Training and the attentional blink: limits overcome or expectations raised? *Psychonomic Bulletin & Review* **21,** 406–411 (2014).
55. Oostenveld, R. & Praamstra, P. The five percent electrode system for high-resolution EEG and ERP measurements. *Clinical Neurophysiology* **112,** 713–719 (2001).
56. Keil, A. *et al.* Committee report: Publication guidelines and recommendations for studies using electroencephalography and magnetoencephalography. *Psychophysiology* **51,** 1–21 (2014).
57. Bigdely-Shamlo, N., Mullen, T., Kothe, C., Su, K.-M. & Robbins, K. A. The PREP pipeline: standardized preprocessing for large-scale EEG analysis. *Front. Neuroinform.* **9,** B153 (2015).
58. Delorme, A. & Makeig, S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J Neurosci Meth* **134,** 9–21 (2004).
59. Chaumon, M., Bishop, D. V. M. & Busch, N. A. A practical guide to the selection of independent components of the electroencephalogram for artifact correction. *J Neurosci Meth* **250,** 47–63 (2015).
60. Cheadle, S. *et al.* Adaptive Gain Control during Human Perceptual Choice. *Neuron* **81,** 1429–1441 (2014).
61. Kok, P., Mostert, P. & de Lange, F. P. Prior expectations induce prestimulus sensory templates. *Proc Natl Acad Sci USA* **26,** 201705652 (2017).
62. Blankertz, B., Lemm, S., Treder, M., Haufe, S. & Müller, K.-R. Single-trial analysis and classification of ERP components — A tutorial. *NeuroImage* **56,** 814–825 (2011).
63. Ledoit, O. & Wolf, M. A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis* **88,** 365–411 (2004).
64. Sprague, T. C., Boynton, G. M. & Serences, J. T. Inverted encoding models estimate sensible channel responses for sensible models. *bioRxiv* **91,** 642710 (2019).
65. Sprague, T. C. *et al.* Inverted Encoding Models Assay Population-Level Stimulus Representations, Not Single-Unit Neural Tuning. *eNeuro* **5,** ENEURO.0098–18.2018 (2018).
66. Oostenveld, R., Martin, Maris, E. & Schoffelen, J.-M. FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Computational Intelligence and Neuroscience* **2011,** 1–9 (2010).
67. Wolff, M. J., Jochim, J., Akyürek, E. G. & Stokes, M. G. Dynamic hidden states underlying working-memory-guided behavior. **20,** 864–871 (2017).
68. Maris, E. & Oostenveld, R. Nonparametric statistical testing of EEG- and MEG-data. *J Neurosci Meth* **164,** 177–190 (2007).